Chapter 3

Feature Level Fusion Using Canonical Correlation Analysis

3.1 Introduction

This chapter provides the details about feature fusion based data-driven models for EMG application. As mentioned in the previous chapter (Ch.1, Section 1.3.1), feature fusion is considered to be the most effective in decision models, in which data itself or features extracted from multiple input features are fused to reduce the variance in the feature registration process and to span the information space. Such low order feature space always has much robustness in real-life applications, especially when the best feature sets are unknown [12]. Thus, in present learning scenario, feature fusion is becoming one of the most effective ways of large-scale information embedding. Most of the real world bio-medical inference system relies on evidences of multiple signals acquired during investigation. This is due to the typical nature of complex biomedical signal that usually make trivial single-set-based inferences. This chapter explicitly explores the multi-view feature fusion for deriving more reliable decision model.

Decision models involve various steps including pre-processing, optimization and parameter settings. Most important step is the initial information processing step and so, more sophisticated signal processing means are needed in order to improve the performance of the model. In pattern recognition paradigm, over the decades, dramatic improvements have been made in feature extraction and classifiers. Various state-of-the-art feature extraction methods have been developed or modified and applied to decision models, which show steady improvement in this era (Ch.1, section 1.3.5). Still, there exist many shortcomings and challenges which are yet to be addressed and overcome. It is worth noting that decision outputs highly depend on the nature and quality of input feature rather than the choice of classification models [92]. In addressing this issue, the theoretical models have been developed and subsequently, various steps involved

are narrated in an organized way. At the end of this chapter, the performances of the proposed model with various formulated information criteria, are investigated and compared with various state-of-art-methods to highlight the efficacy of the proposed model.

3.2 More informative features

The study aims at incorporating large-volume signal(s) or feature information into the initial framework of learning for obtaining more discriminant physiological or statistical features, which are to be fused and embedded to the learning models. Our proposal is to create multiple input features or multi-view features (MV) either using single or multiple signals associated with a study group. It is due to fact that MVL relies on MV so that it should have most physiologically relevant information and energy contents. Our rationale is, therefore, that the MV covers large-volume information compared to the specific signal feature and use of such feature using proposed method could presumably avoid the feature biasing in the learning framework. It has the potential to be the basis of the CCA-based learning.

Each time-window frame of specific signal (considering signal is the sum of multiple time-window frames) is dissimilar in nature and contains different degree of information. Therefore, first hypothesis induced herein, is that the MV derived from multiple frames of a signal should have most physiologically relevant information and energy contents about the underlying process. The second hypothesis is more generalized version of the former and it aims to formulate the MVs using multiple signals associated with a particular study group. The nature of signals depends on disease profiles, subgroup of disease and age of subject and therefore, specific pre-defined may not provide usable information to diagnose the disorders. As a consequence, learning from such MVs could provide a more robust and effective solution for pattern recognition tasks [1, 9, 45, 48, 49, 52]. This study focuses two aforementioned hypotheses in sequential order (Case I and II analysis) and explore the subsequent improvement of model performance with later over the former hypothesis.

Hypotheses employ statistically independent MV using DWT in conjunction with directly evaluated MV. Unlike DWT methods, the proposed method used wavelet only to transform input space and to evaluate domain independent MV. To the best of author knowledge, it has not been used earlier in the context of quality improvement via feature fusion technique. The reason for using DWT based MV in our method is that synchronization of multi-domain features improves the generalization ability of input space [7]. The analysis is carried out in MATLAB (The MathWorks, Inc., Natick, United States) on an Intel (R) Xeon (R) machine (Precision T3500) with processor 2.8 GHz and 8 GB of RAM.

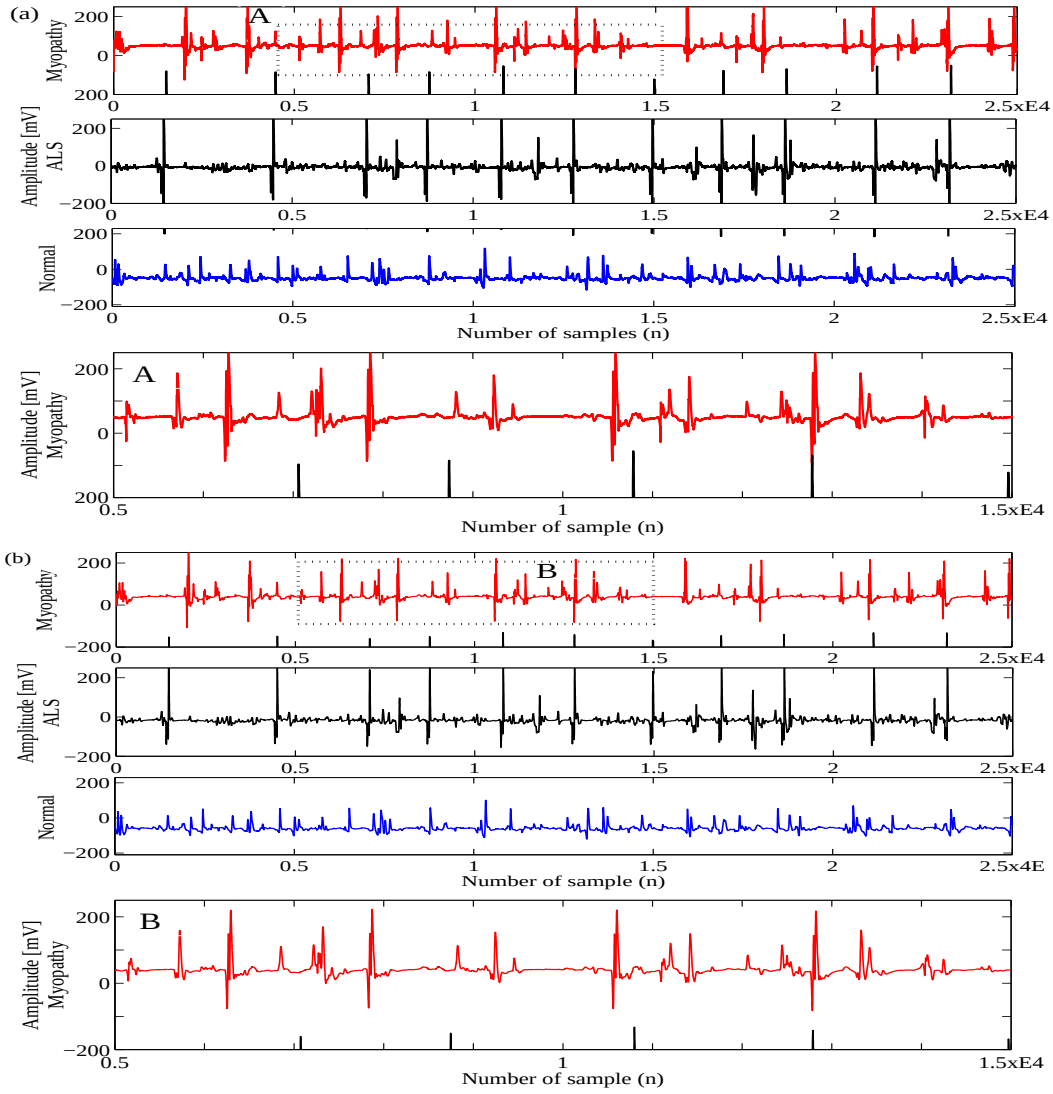

Fig. 3-1: (a) Indicates three typical EMG patterns after scaling, i) myopathy (top), ii) ALS (middle) and iii) normal (bottom), while corresponding filter signals are depicted in (b). Note that $c = 9$ is normalized parameter. This signal processing steps improves SNR to $\approx 18\%$. Zoom subfigures A and B of dotted portions in (a) and (b) indicate clarity of original and filtered signals.

3.3 Formulation of mult-view features: Case I analysis (MVF-Case-I)

Typically users are interested in extracting knowledge from the signals to be analyzed, however, recording signals often contaminate with noises. Therefore, time-course signal can be categorized into two parts-signals that carry valuable information of phenomena of interest and noise or interference (unwanted components). It can be expressed as follows:

$$\mathbf{Y}[\mathbf{n}] = \mathbf{y}[\mathbf{n}] + \eta[\mathbf{n}] \tag{3.1}$$

$\mathbf{y}[\mathbf{n}]$ and $\eta[\mathbf{n}]$, $n = 1, \ldots, N$, indicate signal and noise contribution respectively. It is, thus, essential to remove the artifacts from the signals before analysis. Signal processing including filtering, transformations significantly helps detaching out such components

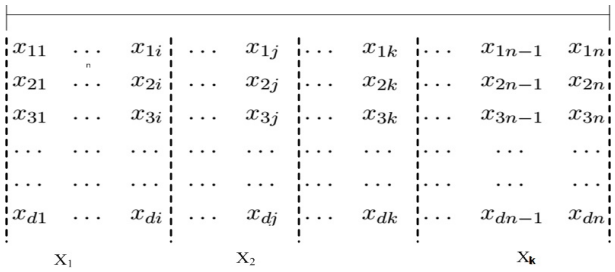

Fig. 3-2: Decomposition of signals and generation of multi-dimensional features using uniform signals sequences.

from the signals and makes feasible the analysis. However, for unknown frequency distributions of signal as well as noise, an appropriate filter setting becomes difficult. In such case, CCA-based method provides an effective solution.

The most energy contents of biomedical signals usually falls in the lower range of frequency axis. As mentioned earlier (See, Ch.1, Section 2.4) for EMG signals, it falls in the range of 0-1 kHz and its dominate energy concentrates in the range of 20-500 Hz. The signals are filtered using a twenty-order Kaiser window based low pass filter with cut off frequency of 0.5 kHz. Then these signals are re-sampled to an even number for ease of analysis. It is seen that the signals have considerably high amplitude in terms of μV and thereby, transformed them using $\mathbf{X}[.] = \mathbf{Y}[\mathbf{n}]/c \pm b$, where c is user define constant, can have any value in the range $0 < c \leq max(||Y[n]||)$ and b is the scaling constant. Fig. 3-1 shows raw typical EMG signals and filtered signals. The filtered signal is uniformly decomposed into a set of sequences [113] and then, taking equal number of sequences, multi-dimensional feature matrices are formulated as-

$$\mathbf{X} = [X_1(i,j), X_2(i,j), \ldots, X_k(i,j)], X_n \in \mathbb{R}^{k \times n_i} \tag{3.2}$$

Here $n_i = n/L$, L is the number of sequences and each row in $X_k(i.,)$ represents the consecutive signal sequences. Understanding of multi-dimensional feature as outlined in Eq.(3.2) and Fig. 3-2. Symmetrical nature of decomposed signal sequences are visually investigated using EMGLAB [114], which help visual investigation of dominated MUAPs in a particular signal sequence, as well as ensures whether consecutive features retain the morphological symmetry in terms of dominated MUAP or satellite MUAP or not. It is important since the existence of correlation between pair of input and subsequent low order features based on correlation can only provide key characteristic of high-dimensional input features.

3.3.1 Canonical correlation analysis

CCA is widely used multi-data processing methods to analyze the mutual relationships between two sets of variables [26, 27]. It finds two sets of low order projected vectors,

one for each view, such that they are maximally correlated in mapping space, that can be utilized to efficiently characterize the object.

Let us assume two multi-view feature matrices X_1 and X_2 as defined in Eq.(3.2). However, for mathematical simplicity they are represented by X and Y. Therefore, PCA is applied to the feature matrices [6]. As reported in [6], it can further avoid encountering the singularity issue that occurs in many real-world small-size-problems. Then, mean of each row from the PCA reduced matrices are removed to make centered data matrices. Let us define two linear transformations, also known as canonical variates of feature matrices X and Y^* as follows:

$$\left.\begin{aligned} u &= A_{x_1}x_1 + \cdots + A_{x_k}x_k = A_x^T X \\ v &= B_{y_1}y_1 + \cdots + B_{y_k}y_k = B_y^T Y \end{aligned}\right\} \tag{3.3}$$

CCA finds weight vectors $A_x = [A_1, \ldots, A_q] \in \mathbb{R}^{d\times P}$ and $B_y = [B_1, \ldots, B_q] \in \mathbb{R}^{d\times Q}$ that maximize the correlation ρ between the variate u and v by solving following optimization problem [44].

$$\max_{A_x, B_y} \rho(u, v) = \frac{E[uv]}{E[u^2]E[v^2]} \longrightarrow \frac{A_x^T \Sigma_{xy} B_y}{\sqrt{(A_x^T \Sigma_{xx} A_x)(B_y^T \Sigma_{yy} B_y)}} \tag{3.4}$$

where Σ_{xx} and Σ_{yy} are autocovariance matrices. Σ_{xy} and $E[.]$ are cross-covariance matrix of X and Y and mean respectively (Note that $\Sigma_{xx} = \Sigma_{yy}^T$)[†]. The overall covariance matrix C that includes Σ_{xx}, Σ_{xy}, Σ_{yx} and Σ_{yy}, cover all feature information on their associations. Optimization problem Eq.(3.4) is solved by using Lagrange multipliers subjecting to $A^T\Sigma_{xx}A = B^T\Sigma_{yy}B = 1$. That means solving of this optimization problem requires to solve following standard eigenvalue equations.

$$XY^T(YY^T)^{-1}YX^TA = \alpha^2 XX^TA, \tag{3.5a}$$

$$YX^T(XX)^{-1}XYB = \alpha^2 YY^TB. \tag{3.5b}$$

Here α^2 is the diagonal matrix or square of canonical correlation with d non-zero elements. The solution gives a set of features pair corresponding to d-correlations which are in descending orders, i.e., $\alpha_1 \geq \alpha_2, \ldots, \geq \alpha_d$. The canonical correlations indicate how close the projected vectors are in orthogonal subspaces as shown in Fig. 3-3. The higher value of diagonal elements indicates close proximity of vectors of two subspace. Intuitively, first few pairs show significant proximate behavior, i.e., they well-capture the particular information from pair variable. Thus, the high dimensional feature is well confined to a low-order structure that preserves most of the energy contents of the sets. Moreover, to avoid over-fitting of model and singularity of both scatter matrices, the algorithm is implemented with regularization parameters α and β[‡] [115]. Nonethe-

*For inputs $X \in \mathbb{R}^{k\times P}$ and $Y \in \mathbb{R}^{k\times Q}$, CCA limits the feature dimension to the ranks of input variables, i.e., $d \leq k = rank(X, Y)$ $(n_i = Q, P)$

†$\Sigma_{xx} = XX^T$, $\Sigma_{yy} = YY^T$, $\Sigma_{xy} = XY^T$ and $\Sigma_{yx} = YX^T$.

‡Although no theoretical guidelines exist for proper setting of regularization parameters. However,

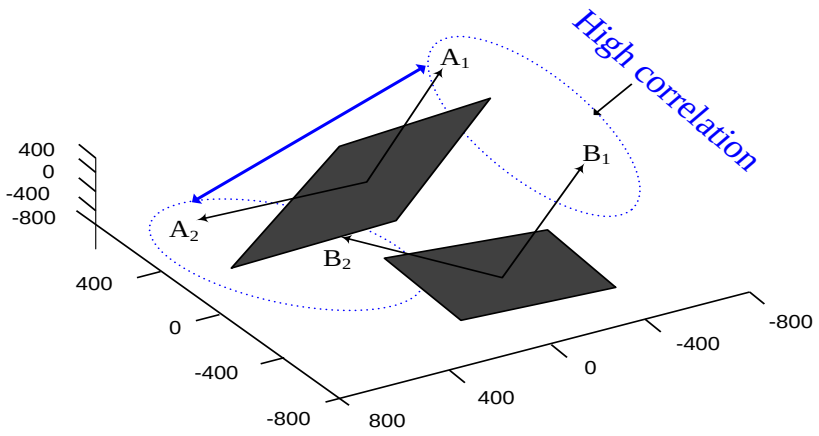

Fig. 3-3: Linear subspace of two sets of vectors. Correlation exists only between same indices pairs vector, i.e., A_1,B_1, while it is zero for different indices pairs, i.e., A_1, B_2 due to orthogonality [21]. Dotted pairs have high correlation.

less, it does not transform the projections as Eq.(3.4) is independent of scaling to A_x and B_y. Under regularization condition, the aforementioned Eq.(3.5a)-Eq.(3.5b) can be generalized as-

$$XY^T(Z_2Z_2^T)^{-1}YX^TA = \alpha^2 Z_1Z_1^TA, \tag{3.6a}$$

$$YX^T(Z_1Z_1)^{-1}XY^TB = \alpha^2 Z_2Z_2^TB. \tag{3.6b}$$

Where $Z_1Z_1^T = \Sigma_{xx} + \alpha I$ and $Z_2Z_2^T = \Sigma_{yy} + \beta I$ and I is identity matrix of same dimension of Σ_{xx} and Σ_{yy}. Use of α and β improves the regularization of problem and also avoids the over-fitting of model. It is worth mentioning that dimensionality, as well as the noise level in the feature dataset, are usually high that arises during signal acquisition or due to human activities. Furthermore, the feature vectors corresponding to low order correlation are insignificant which also account for noise contribution and not considered in this analysis. The aforesaid optimization problem can be solved by using efficient singular value decomposition (SVD) technique [21]. Assume that X and Y form unitary orthogonal bases for two linear subspaces (Fig. 3-3) and according to the theory, the SVD of $X^TY \in R^{d\times d}$ be-

$$X^TY = U\Lambda V^T = [U_1, U_2]D[V_1, V_2]^T = U_1\Lambda_d V_1^T, \tag{3.7a}$$

$$A_x = XU_1, B_y = YV_1. \tag{3.7b}$$

where U and V are two left and right singular orthogonal matrices of Σ_{xy} and Σ_{yx} *respectively*[§] Also $X^TU_2 = 0$ which infers that U_2 falls in the null space of X, i.e., uncorrelated component [115]. The diagonal elements in D are in descending order

moderate parameter values are effective to maintain the balance of learning model and thereby we set $\alpha = \beta = 0.5$ empirically.

[§] $D = diag(\Lambda_d, 0)$ is the diagonal matrix. According to the CCA, correlation exists only in-between same indices pair vectors, say $\{A_1, B_1\}$ and zero correlation in-between different indices pairs, say $\{A_1, B_2\}$. Therefore, non-diagonal elements in D are zeros.

and measure the canonical correlations between a pair of projected vectors given by Eq.(3.7b). By reducing the feature dimensionality based on the degree of similarity index between two input features, the learning parameters can be further reduced, which in turn reduces the computational cost. Such strategy promises to provide finer interpretation of problems and a good breakthrough in many aspects of signal processing era. Detail mathematical solution and proof of both regularized and non-regularized problems using eigen-decomposition technique are given in [21].

3.3.2 Variability and stability analysis

This section aims to provide some ideas, perspectives, guidelines and key direction to transform high dimensional feature into low order features based on measurement of natural changes in modality signals. Thus, there is an obvious interest in signal variability and stability analysis. A well-accepted framework in statistical paradigm is adopted for providing the striking evidence to the variations and consequent stability [116,117]. The feature extraction and selection (FES) is directly related to signal variability, i.e., change in morphology with time, shape and size etc. The obtained features can retain the most latent information to be used for automated decision models for correct interpretation with low computation burden [118].

The signal variability can be either intrinsic or extrinsic. In intrinsic variability, a major endeavor in the biomedical field is understanding the changes in modality signals. It is well-known that the nature of signal obtaining from the same modality the diverse character for various study patterns. Even within same study pattern, there are some inherent variability due to the dependency of signals on number of factors. Furthermore, variability of data from specific time-window to the next are also prominent. Therefore, interpretation of measurement based on specific signals or frames might be difficult. In this context, MVL have been proven to be successful in covering a wide range of modalities data. Beyond the strengths of MVL with respect to others, the parameter degeneracy and constraints are two major ambiguities. A systematic approach to interact with modality data and understand the nature of changes is adopted using the CCA. It finds reasonable statistics for fluctuation measurement and subsequently allows to reduce the allowable uncertainties volume of features.

The basic mathematical preliminaries to provide a more accurate meaning to this ideas are described here. The model in Eq.(3.2) provides a set of consecutive features which can be expressed as follows:

$$F(x) = F(X_1, \ldots, X_k) \tag{3.8}$$

The consecutive pairs $\{(X_1, X_2) \ldots (X_{k-1}, X_k)\}$ are applied for correlation evaluations, based on which variations as well as stability of feature or signals are examined. It

is referred to as *variability within-subjective signals* (VWS). The local variation, i.e., feature to feature variation, can be uniquely characterized from a correlation index r, i.e. correlation matrix as shown in Eq.(3.9). The parameter values (i.e., r) are estimated for all possible pairs and presented in matrix form, referred to as collective correlation coefficient matrix (CCCM), where each row indicates the correlation profiles between two sets of projected features of input features in transformed subspace. The number of parameter value is limited by $d = rank(X_{k-1}, X_k)$. The correlation matrix r is given as follows:

$$r = \begin{pmatrix} r_1 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \dots & r_d & 0 \\ 0 & 0 & \dots & r_i \end{pmatrix} = Diag(r_1, \dots, r_d, r_i) \tag{3.9}$$

Where r_i indicates the correction between pair of vector of same indices. To enrich the analysis and understand the mechanism, graphical and overall correlations (OC) are evaluated. Graphical analysis reveals the variations of local pair input features, while the OCs indicate the overall quantitative deviation of pairs. Thus, we have looked at k-features and for $k/2$ feature pair, k being even, the OC is evaluated.

$$\Gamma_{k/2} = \sum_{i=1}^{d} r_i^2 \tag{3.10}$$

Here $\Gamma_{k/2}$ indicates the OC. Any change in $\Gamma_{k/2}$, i.e., $\Delta = \Gamma_1 - \Gamma_2$ indicates close similarity between two input feature vectors and lesser variation between two. This indicates higher stability of features. This analysis is particularly carried out to evaluate an appropriate feature dimension by avoiding unnecessary variance of feature space due to unwanted features.

In certain scenario, the variations of signals are in terms of temporal energy profiles (i.e., rms value) extracting from time-frame of signals obtained using decomposition method [58]. One possible difficulty with this approach is the choice of feature dimensionality. Instead, a specific set of domain features were considered for analysis. Unlike this, the adopted approach provides a well-define framework to draw the conclusions as stated below.

i). It elaborates the latent information variation and its extension within same study group patterns.

ii). A set of low order features corresponding to each set of correlation can be obtained, which can be further reduced by imposing correlation threshold.

iii). It gives rise to realistic model with the proper guideline that enables true interaction with the heterogeneous dataset which is not pre-known or sufficiently

understood.

iv). CCA finds mutual information between features ignoring uncorrelated information.

For extrinsic variability study, the signals associated with subjects under same study group which are often diverse, are considered. The acquisition setup provides recordings at different resolution and further, it relies on recording sites, type of muscle, force recruit to muscle (i.e., in case of EMG). Therefore, the uncertainty associated with various signals might be different. In this case, information registration in decision models relies on correlation or interaction among multiple signals belonging to various subgroups of same study subject which is the most prevalent challenge in this scenario. For this, we devised our experiment extracting feature sets of same order similar to Eq.(3.8) for each subject separately and subsequently, a feature vector from one set X_k is coupled with corresponding feature of other set Y_k, i.e., $\{(X_1, Y_1) \dots (X_k, Y_k)\}$. It is referred to as the *variability in intra-subjective signals* (VIS). It would provide the deviation of features of subgroup to next, from where signal variation from subgroup to subrgoup under same study group could be concluded. In the proceeding section the FES is widely explored based on VWS and VIS.

3.3.3 Feature extraction and selection (FES)

The most crucial endeavor to be undertaken in decision models is the FES. The term *feature extraction* indicates the evaluation process of features from the input space using mathematical transformation, while *selection* stands for the *choice* of appropriate features from transformed domain that reflects the underlying phenomena [119]. So, FES need to be an efficient, sensitive and intuitive so as to actively communicate between two systems [1]. Often human play active roles in the choice of proper strategy or model and also in the way it is used in practice. The latter implies that the design of the model and FES procedure rely not only on theoretical scheme but also on suitability and adaptability to the environment preferences.

Many well-known FES methods (decomposition and extraction) were employed in EMG classification models (See, Ch. 2, Section 1.3.5, [60]). However, due to very heterogeneous and complex nature of signals, many methods fail in extracting useful information from signals. Work on various FES methods are in progress to incorporate time-frequency, DWT [120, 121] or statistical features [58], however, prior to many recent methods, e.g., [23], other methods such as TD, AR, AR+RMS and AR+RMS+TD[¶] are also remarkable. Multi-DWT is often common in non-stationary signals [122, 123]. Despite massive work over the decades, very few methods are useful in practical applications, mainly due to inappropriate choice of FES method. The existing FES methods have the limitations such as-

[¶]AR: Autoregressive coefficient, RMS: root mean square and TD: Time domain.

i). Involvement of large number of assumptions, wherein some are unrealistic in nature to the phenomena which results in undesirable outcomes [13].

ii). Excessive number of constraints and the high degree of freedom associated with various models make them difficult in using real-life scenarios without the support of expert personnel.

iii). Lack of explanation for reason of higher model performance, consideration of feature searched space, and reliability of methods, complexities etc.

Therefore, this section addresses a theoretical data-driven approach undertaking the aforementioned perspectives. It is restricted to a three class EMG classification problem. Within this, we adopted two-stage approaches; in the first case, CCA transformation is applied in pairwise manner over all possible pairs of MV and subsequently evaluated the transformed features for all pairs, and in the second case, DWT-based MVs are evaluated and then subjected them for analysis. As discussed in the previous section (See, Section 3.2), the choice of statistically independent features and subsequent fused features improve the generalization ability of feature space.

The merits of the DWT have been discussed in Ch.1, Section 1.3.6. The DWT transforms the signal $x[n]$ of length N through a high-pass filter (H) and low-pass filter (L) with impulses $h[n]$ and $g[n]$ respectively. In first level decomposition, filter outputs give $d^1[n]$ and $a^1[n]$, which are known as detail and approximation. The output a^1 is further downsampled by 2 and pass through second set of filters as in Fig. 3-4. The second level decomposition constitutes two frequency subband components similar to previous set. These are mathematically expressed as-

$$d^2[n] = \sum_{k=0}^{N/2-1} a^1[k]h[n-k], \tag{3.11a}$$

$$a^2[n] = \sum_{k=0}^{N/2-1} a^1[k]g[n-k]. \tag{3.11b}$$

An appropriate choice of prototype or mother wavelet function that accurately fits the application and signal is essential. This function determines the coefficients of high-pass and low-pass filter of the DWT. As shown in [58], $db2$ of the Daubechies wavelet function is efficient for EMG analysis. It further suggested that second level of decomposition and use of coefficients are reasonable. It is proposed to use localized low-frequency components for deriving the features since most energy contents of biomedical signals, specifically EMG and EEG, usually fall in low frequency range (e.g., 0-1 kHz for EMG). The higher level of decomposition further narrow downs the localization of frequency in subbands coefficients. The a^2 components of all selected signals are used as input signals to find DWT-MVs. Fig. 3-5 shows the approximate and detail coefficients for three groups of subjects indicating DWT signal selection mechanism.

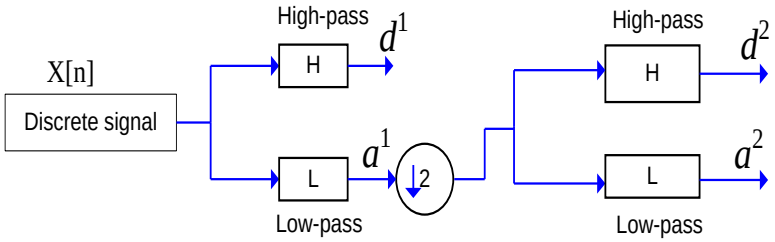

Fig. 3-4: Two-level decomposition tree of a discrete signal of the DWT.

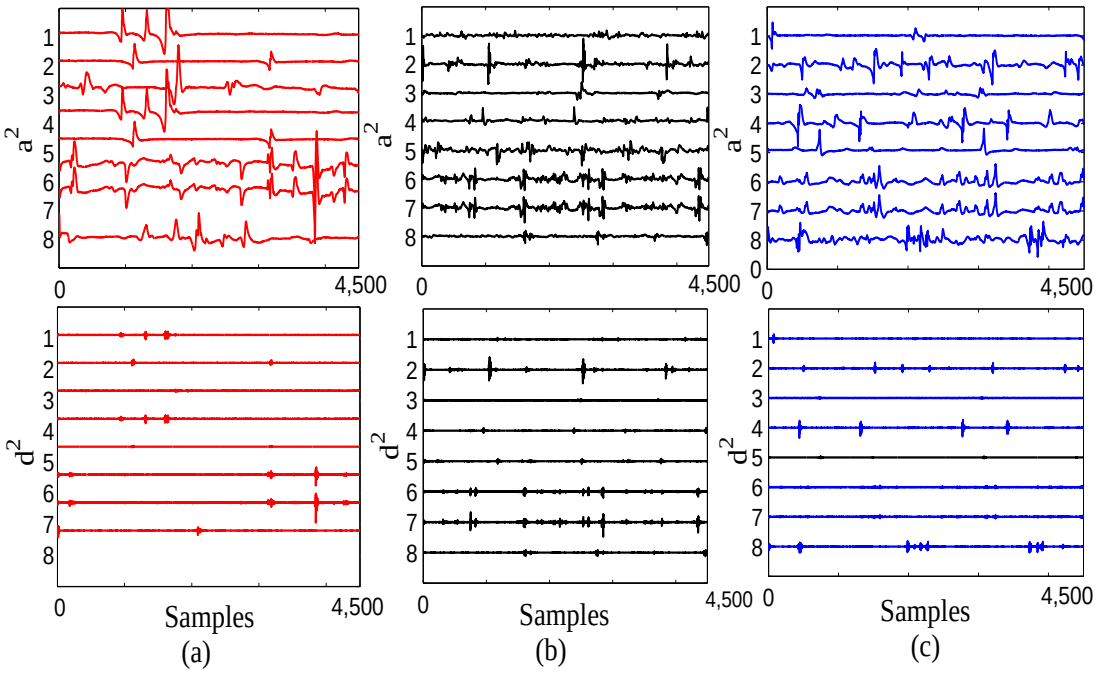

Fig. 3-5: Second order wavelet approximate (a^2) and detailed coefficients (d^2) used to formulate domain independent multi-view features for (a) ALS, (b) myopathy and (c) normal.

Pattern recognition algorithms use both local and global features extracted from modality signals. The direct approach employs features from EMG signal-frame, while MUAP-based method uses local or dominated MUAP features [58]. However, many MUAPs are non-stationary and finding dominated MUAP from IP EMG is also difficult task, which in fact requires domain expert. Albeit good results, MUAP based approaches may not be so transparent to the clinical practice. Further, due to the *within-signal variation* (intrinsic) and *intra-signal variation* (extrinsic), feature extracted from specific time-window frame may fail in representing the underlying process. The proposed model finds two set of features using independent input pair feature. FES module as shown in Fig. 3-6 describes feature extraction scheme.

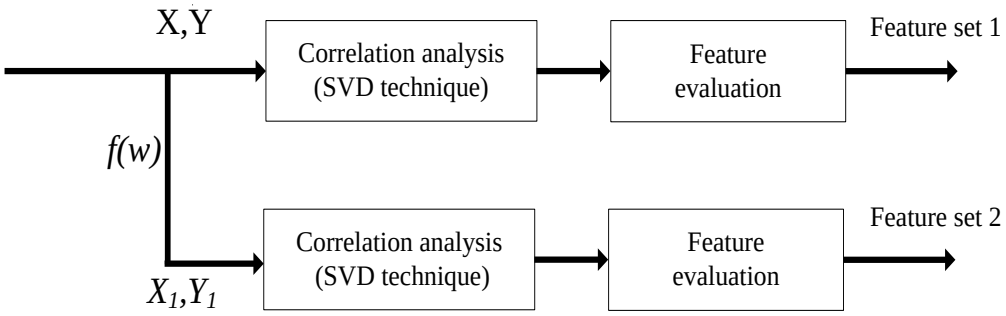

Fig. 3-6: Diagram showing evaluation of two feature sets using each direct MV pair and DWT-MV pairs. Two sets are statistically independent. $f(w)$ indicates DWT transformation applied over direct features to find DWT-MV.

3.3.4 Feature fusion and classification strategy

Data fusion or feature fusion provides more comprehensive solutions for many real-world problems (See, [124–126] and references therein). It receives wide attention in many applications [6, 11] due to inherent feature extraction strategy using large-volume data. Luo *et al.* in [127, 128] introduced various levels of fusion and suitability of feature fusion for data classification tasks. Followings are essential merits of feature fusion.

i). It create the framework that can handle large volume data available in many real-worl problems.

ii). It enables interaction of multiple measurements or features (i.e., multi-view features) of same or different modalities profiles.

iii). It provides low order discriminant features that convey more information of underlying phenomena [6], which are more suitable for machine learning applications.

The challenges inherent in many application domains [1] (and others) and in fusion technology [124, 125], it is clear that the feature extraction and their fusion at an appropriate level would completely describe a given phenomena and would be relevant to diagnosis applications.

Alternative to the feature concatenation, we build a global fusion model using two proactive-concatenation and summation techniques. The scheme as shown in Fig. 3-6 using model Eq.(3.7b) evaluates two sets of transformed features vectors, termed as projected vectors (PV) or feature vectors (FV). The discriminate features are fused using concatenation and summation techniques respectively, as follows:

$$\phi_i = Z_i^m = \begin{bmatrix} X^\star \\ Y^\star \end{bmatrix} = diag \begin{bmatrix} A_x & B_y \end{bmatrix}^T \begin{bmatrix} X \\ Y \end{bmatrix}, \tag{3.12a}$$

$$\psi_j = Z_j^m = X^\star + Y^\star = \begin{bmatrix} A_x \\ B_y \end{bmatrix}^T \begin{bmatrix} X \\ Y \end{bmatrix}, \tag{3.12b}$$

As defined in the section 3.3.1, A_x and B_y are weight vectors, and $Z^m_{i(j)}$ is known as local canonical correlation discriminant features (CCDF) of $i(j)^{th}$ subgroup (note that $m = k/2$). In order to cover global features, new fusion models DMI and DMII with single level fusion are evaluated. In other word, it gives single-domain feature fusion, refer to as $F = 1$.

$$(DMI, F = 1)\Phi_i = [Z^1_i, \ldots, Z^m_i], \tag{3.13a}$$

$$(DMII, F = 1)\Psi_j = [Z^1_j, \ldots, Z^m_j]. \tag{3.13b}$$

Dimension of fusion models relies on order selection of feature vectors A_x and B_y. However, the local features $Z^m_{i(j)}$ is maintained at constant order to avoid difficulty in combining them. For low dimensional case, the global features are obtained by juxtaposing local CCDFs as in Eq.(3.13a)-(3.13b) and then used feature reduction technique. In case of high dimensional CCDFs, mean of all local feature provides an appropriate solution. It is worth mentioning that the summation techniques, Eq.(3.12b) and Eq.(3.13b) (i.e., ψ_j and Ψ_j) is superior to the concatenation techniques, Eq.(3.12a) and Eq.(3.13a) (i.e., ϕ_i and Φ_i) due to low dimensional feature combinations. Thus, one can uniquely define the low order model to comprehensively represent the discriminate information extracting from input features via transformations. These strategies waive theoretical bottlenecks, computational complexity and reduces time. However, the pros and cons of various combination of features in applying them in real-time applications are discussed in the context of EMG signal classification in the proceeding sections.

As discussed in previous section, the advocated scheme attempts to combine domain independent features to improve the generalized ability of learning model. Following this principle, the domain independent features are fused similar to Eq.(3.12a)-(3.12b) and then combined models are evaluated similar to Eq.(3.13a)-(3.13b) for $F = 2$, as follows:

$$Z^m_i = \begin{bmatrix} X^\star \\ \vdots \\ Y^\star_1 \end{bmatrix} = diag\begin{bmatrix} A_x & \ldots & D_y \end{bmatrix}^T \begin{bmatrix} X \\ \vdots \\ Y^f_2 \end{bmatrix}, \tag{3.14a}$$

$$Z'^m_j = X^\star + \cdots + X^\star_1 + Y^\star_1 = \begin{bmatrix} A_x \\ \vdots \\ D_y \end{bmatrix}^T \begin{bmatrix} X \\ \vdots \\ Y^f_1 \end{bmatrix}, \tag{3.14b}$$

$$(DMI, F = 2)\Phi'_i = [Z'^1_i, \ldots, Z'^m_j], \tag{3.14c}$$

$$((DMII, F = 2)\Psi'_j = [Z'^1_j, \ldots, Z'^m_j]. \tag{3.14d}$$

Here, X_1 and Y_1 are wavelet multi-view features of X and Y respectively. These fusions are referred to as multi-domain fusions and obtained discriminant features are known as global generalized CCDF (gCCDF). In evaluating the appropriate dimensional features, DMI and DMII (for $F = 2$) are subjected to LDA analysis and transformed features are

evaluated based on LDA selection criteria.

3.4 Formulation of mult-view features: Case II analysis (MVF-Case-II)

According to the second hypothesis as mentioned in the Section 3.2, this section provides detail description of MV formulation unlike to the case I analysis and feature fusion scheme to diagnose neuromuscular disorders. It is purely assumption free data-driven based scheme. The core focus of this analysis is to extract features from more generalized MV inputs and to fuse and embed to the classification models. It also focuses on the effectiveness of the algorithm in the context of results obtained over various EMG datasets and comparison with state-of-the-art-methods. Under this circumstance, two high dimensional feature formulation strategies S-I and S-II are introduced, which will be used independently to evaluate low order features for model learning. This scheme is referred to as multi-view CCA (mCCA).

3.4.1 Strategy I (S-I)

A given dataset contains C subject groups and each group is divided into c subgroups. Each subgroup consists of n subjects and employing q signals from each subgroup an MV X_c is evaluated as shown in Fig. 3.7(a). In this way, we obtain the set of X_c, referred to as X_C to represent the subject group C. In order to find statistically independent features, DWT is performed over each signal of X_c and using low-frequency components independent feature matrix $X_{c\omega}$ and subsequent set $X_{C\Omega}$ are evaluated as-

$$X_c = [x_1, \ldots, x_q]^T \in \mathbb{R}^{q \times p} \tag{3.15a}$$

$$X_{c\omega} = [\tilde{x_1}, \ldots, \tilde{x_q}]^T \in \mathbb{R}^{q \times \tilde{p}} \tag{3.15b}$$

$$X_C = \{X_1, \ldots, X_c\}; X_{C\Omega} = \{X_{1\omega}, \ldots, X_{c\omega}\}. \tag{3.15c}$$

Where x_q in Eq.(3.15a) is the one-dimensional p samples and $\tilde{x_q}$ in Eq.(3.15b), is the corresponding low frequency wavelet components. Two MV sets in Eq.(3.15c) are statistically independent and the feature obtained from them by using effective way improve the recognition performance. Each MV and its sample delayed version is highly correlated [44]. Therefore, this pair-wise strategy is used for all possible pairs of two sets in CCA transformation.

Table 3.1: Single-domain MV and sMV according to the formulated S-I and S-II

Subject-groups [C]	Subgroup MV [for c]	Sets of MV [S-I]	Total sMV (Pair) [S-II]
ALS	X_1, X_2, X_3 (3)	X_C (1)	3×8 = 24 (12)
Myopathy	X_1, X_2, X_3 (3)	X_C (1)	3×8 = 24 (12)
Normal	X_1, X_2, X_3 (3)	X_C (1)	3×8 = 24 (12)
C= 3	Total c= 3+3+3	Total= 3	Total = 72 (36)

3.4.2 Strategy II (S-II)

In S-II, MVs in Eq.(3.15a)-(3.15b) are decomposed uniformly. Each signal is segmented into a number of sequences with equal number of samples. Signal sequence and corresponding sequences of other signals are embedded to template sub-MV (sMV) as shown in Fig. 3.7(b). Similar to S-I, two independent sets of sMV are formulated as-

$$\tilde{X}_c = \{\tilde{X}_1, \ldots, \tilde{X}_r\}; \tilde{X}_{c\omega} = \{\tilde{X}_{1\omega}, \ldots, \tilde{X}_{r\omega}\} \tag{3.16}$$

Here $\tilde{X}_r$ and $\tilde{X}_{r\omega}$ indicate sMV obtained from the MV. In Eq.(3.15c), two sets of MVs are derived from C while in Eq.(3.16), two new feature sets of sMV are derived from c. The consecutive sMV retain their morphological symmetry owing to same subgroup of unimodal data. Visual examination of signals with EMGLAB and MATLAB show decomposed features retain their symmetry in terms of morphology. In such case, mCCA can extract the relevant feature subsets from the pairs of consecutive sMV based on correlation [21] leading to an efficient model. In S-I, low order features are evaluated from the common subspace of MV and delay version, while in S-II these features are extracted from the common subspace of consecutive sMV.

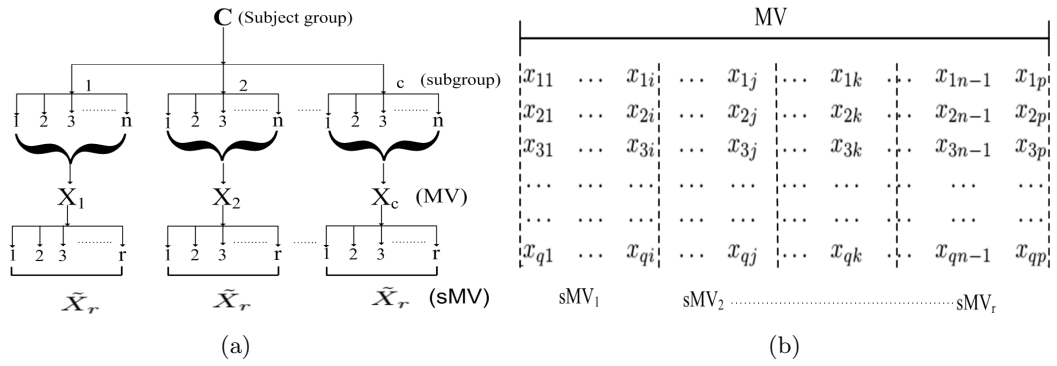

Fig. 3-7: a) Generation of MVs and sMV from c subgroups of C subject groups according to S-I and S-II respectively. Here n indicates number of subject in each subgroup and r indicates the number of sMV from each MV, and b) evaluation of sMV from the MV as shown in Fig. 3.7(b) (a) using uniform decomposition strategy. Each row in MV represents 1D EMG signal.

3.4.3 Feature extraction and reduction

Table 3.1 shows the obtained MVs and sMVs for three study groups (See, Table 2.1) using S-I and S-II. Mathematically, they are expressed as follows:

$$X_C(MV) = [X_1, X_2, X_3]_{i=1,..,C} \tag{3.17}$$

$$X_c(sMV) = [\tilde{X}_1, \ldots, \tilde{X}_r] \tag{3.18}$$

Here $c = 3$ (subgroups) and $C = 3$ (groups). For each group, 3 MVs are evaluated, one for each subgroup using S-I as indicated by Eq.(3.17). In S-II, each MV is divided into 8 sMVs (i.e., r=8) and 24 (=3 × 8) sMV are evaluated for each group (i.e., Eq.(3.18)). Same way, DWT-MV and sMV are evaluated. In finding the MVs, various q[‖] are selected and optimal is estimated.

MVs and its corresponding sample delayed versions are pair-wisely projected to the subspace and then, the features are evaluated using Eq.(3.4). Fig. 3-8(a) shows the correlation over three pairs of MVs in ALS group that reveals the existence of high correlation (i.e., r) between MV and delayed MV which also affirms the evidence of [44]. It is worth noting that number of CCA components is equal to the dimension of input feature which is equal to 8. Here as per analysis, zero correlation corresponding to this number indicates nonexistence of similarity between 8^{th}-pair of feature. This way, three feature sets are estimated for three subject groups corresponding to three correlation profiles. mCCA finds another three feature sets using DWT-MV similar to Eq.(3.17). Thus, for each group, six feature sets (= 2 × 3) are evaluated in S-I.

In S-II, correlations are evaluated for consecutive pairs of sMV. Fig. 3-8(b) shows the within-group sMV-correlation which reveals the suitability of highly correlated features that capture intrinsic information of input features. In evaluating MV and sMV similar to the Table 3.1, second order db2 wavelet function [58] is used and only low frequency components A_n are employed while high frequency components D_n are ditched from analysis. Most user interest energy contents of biosignals fall in lower frequency scale. In this analysis, we obtained optimal signal value at q=8 and evaluate matrices MV and sMV of dimension 8 × 258000 (S-I) and 8 × 32250 (S-II) respectively. Also, correlation threshold is set at 7 and corresponding features are considered for analysis. Thus, the proposed method reduces the feature dimensionality by using correlation threshold and SVD. The work flow of the proposed method is shown in Algorithm 1.

3.4.4 Feature fusion, transformation and classification

Summation technique is more suitable due to dimensionality issue (See, Section 3.3.4). Therefore, similar to Eq.(3.14b), the discriminant feature Z_{ij} for all possible pair input

[‖] The value of $q = 4, 6, 8$ and 10 are chosen in this study

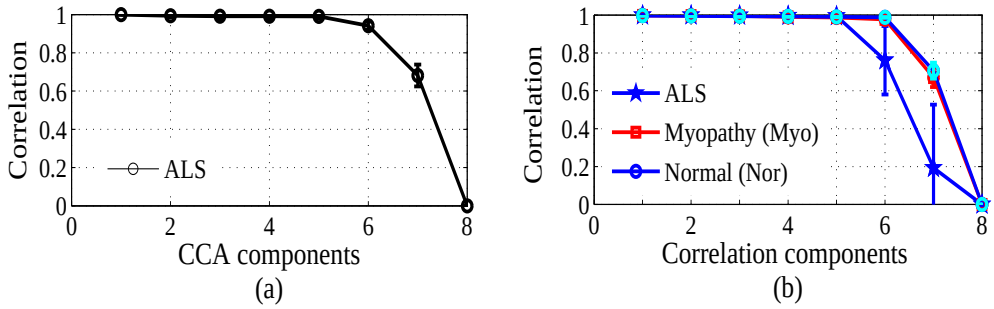

Fig. 3-8: Mean correlation with SD as error bar between a) multi-view feature and its delayed version (ALS), and b) Consecutive sub-multi-view in three study groups.

Algorithm 1: Feature fusion algorithm using S-I and S-II

Require: Select various value of q.

Finds MV and sMV using S-I and S-II
$C \longrightarrow \{1, \ldots, c\}$
$X^{(i)} \in R^{q \times p} \longleftarrow c$
$X^{C(1)} = \{X^{(i)} \in R^{q \times p}\}_{i=1}^{c} \longleftarrow$ C
Repeat 1-3 for $X_{\omega}^{C(1)}$ in DWT
Do SVD for $X^{C(1)}$ $X_{\omega}^{C(1)} \longrightarrow U \Lambda V^T$
Compute $\longrightarrow \{A_x, B_y\}$ & $\{C_{x1}, D_{y1}\}$
Do feature level fusion $\longrightarrow \{Z_{ij}\}$ for each C,
Set transformation $\Phi_i \longrightarrow Z_{ij} + \gamma I$ for $\gamma_c = 0, \gamma_c \neq 0$
Transform $\{\Phi_i\}_{i=1}^{c(C)} \longrightarrow$ LDA
Compute $S_W = \sum_c N_c(\mu_c - \mu)(\mu_c - \mu)^T$;
$S_B = \sum_{x \in C_c}(x - \mu_c)(x - \mu_c)^T$
do optimization $\longrightarrow \dfrac{det(W^T S_B W)}{det(W^T S_W W)}$;
Solve $\longrightarrow S_W^{-1} S_B$;
Find eigenvectors for d largest eigenvalue;

features (i.e., MV or sMV). The simplified version of aforesaid equation is as follows:

$$Z_{ij} = \sum_{t=1}^{4} T_t(u, v) = A_x^T X + B_y^T Y + C_{x1}^T X_1 + D_{y1}^T Y_1. \tag{3.19}$$

where T_t is canonical variate and Z_{ij} is the gCCDF of $\{ij\}th\ pair$**. Many prior successes of multigroup feature fusion based method like [5, 29] motivated to implement such scheme in our study. According to the Table 3.1, three and twelve gCCDFs are evaluated for each study group using S-I and S-II respectively. Then, evaluated mean gCCDFs as global measures, are transformed using the following proposed model.

$$\Phi_i = Z_{ij} + \gamma I. \tag{3.20}$$

** X_1 and Y_1 are transformed DWT MVs of input X and Y; C and D are corresponding weight vectors

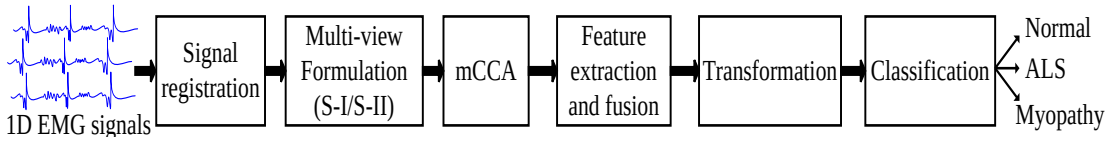

Fig. 3-9: Proposed multi-view feature fusion based scheme for classifying EMG signals to diagnose neuromuscular disorders. Initial signal registration stage involves multiple signals selection from the available resource, processing and orientation of 1D signals into the form of multi-dimensional matrices, followed by subsequent steps as discussed.

where γ and I are class indicator parameter and identity matrix respectively. The dimension of gCCDF was matched with I by padding zeros. This experiment is accomplished with empirical value of $\gamma = 0, 10$, and 20 for normal, myopathy and ALS respectively. It is done to further improve the generalization ability of features by enhancing the separation margin in decision surface [129]. However, the experiment is also carried out without using Eq.(3.20).

As mentioned in Ch.1, Section 1.3.2.1, feature reduction is an essential requirement in data mining and machine learning applications. In order to extract best feature sets, the extracted feature space is further subjected to the LDA, which finds the linear transformation to set the best decision margin among feature clusters. It basically utilizes class-structure information through minimizing within-class variance S_W and maximizing between-class variance S_B [23].

$$S_W = \sum_c N_c (x_i - \mu_c)(x_i - \mu_c)^T \tag{3.21}$$

$$S_B = \sum_{x \in C_c} (\mu_c - \mu)(\mu_c - \mu)^T \tag{3.22}$$

where μ_c, μ and N_c indicate the sample mean of feature vectors, overall mean of the entire sample set and the number of samples in class C respectively.

Fig. 3-9 shows the decision model based on feature fusion strategies S-I and S-II. Use of LDA with k-nn avoids the iterative training and under or overfitting which leads to higher generalization ability of model [23]. k-nn finds the euclidian distance function between the feature in the test set and neighboring healthy and pathological patterns in the training set. The pattern from the test set is categorized based on the class labels of closer patterns.

3.4.5 Performance evaluation markers

To assess the performance, each dataset is partitioned into three subsets, 50% for training, 25% for validation and 25% for testing. The classifier has been supplied with statistical features to be assigned any one of three classes-ALS, myopathy and normal.

To evaluate the model performance, we estimate the sensitivity (Sn), specificity (Sp) and overall accuracy (OA). The OA is given as-

$$OA = \frac{Numberofcorrectclassification}{Totalnumberofclassification} \times 100\% \tag{3.23}$$

Sn, Sp, Ac/OA as well as related parameters-positive predictive value (PPV), negative predictive value (NPV) and error rate (ER) [130–132] demonstrate the balance and efficacy of decision model. These typical parameters are extensively used for better assessment of algorithm performances without going to the analysis. Usually, these parameters are evaluated from the confusion matrix of results. Sn and Sp measure the degree of positive cases (i.e., ALS, myopathy) and negative case (i.e., normal) correctly classified.

$$Ac = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}, Sn = \frac{T_P}{T_P + F_N}, Sp = \frac{T_N}{T_N + F_P} \tag{3.24a}$$

$$PPV = \frac{T_P}{T_P + F_P}, NPV = \frac{T_N}{T_N + F_N}, ER = 1 - Ac. \tag{3.24b}$$

where T_P and T_N are number of positive and negative class identified correctly, F_P is the number of negative class identified incorrectly as positive class and F_N is the number of positive class identified incorrectly as negative class. In multi-label classification problem, i.e., multiple diagnosis, class-specific Sn and Sp are also needed [133].

3.5 Results and discussion

3.5.1 MVF-Case-I

In order to investigate VWS and VIS as discussed in the Section 3.3.2, signals acquired from various muscles under normal or disease profiles as outlined in Table 3.2, are considered. This analysis mainly focuses on demonstrating heterogeneous nature of signals that varies within or in-between intra-subgroup of same group and emphasizing how it limits the dimension of feature space for appropriate realization of underlying information.

3.5.1.1 Variability within-subjective signals (VWS)

Subject groups are partitioned into four different subgroups (e.g., A1-A4) depending on sex, age, duration of disease as well as signal recording sites. However, healthy control subjects have no sign of neuromuscular disorder as such disease duration is marked as "zero"in Table 3.2.

From four observations or signals, six input features (i.e., $k = 6$) are evaluated

Table 3.2: Statistics information of various muscles and study subjects of three groups. Note that durations of normal subject group are zero for all cases.

Subject group	Muscle	a‡	Subgroup	Label
ALS	Abductor polliciesbrevis (AP)	56, 0.5, M	A56MAP0.5	A1
	Tibialis anterior (TA)	35, 5.0, M	A35MT5	A2
	Vastus lateralis (VL)	35, 5.0, M	A35MV5	A3
	Biceps brachii (BB)	67, 1.5, F	A67MF0.5	A4
Myopathy	Tibialis anterior (TA)	28, 12.0, M	M28MT12	M1
	Biceps brachii (BB)	44, 26.0, M	M44MB26	M2
	Biceps brachii (BB)	41, 2.0, F	M41FB2	M3
	Biceps brachii (BB)	26, 1.0, M	M26MB1	M4
Normal	Biceps brachii (BB)	23, 0, M	H23MB0	H1
	Biceps brachii (BB)	26, 0, M	H26MB0	H2
	Biceps brachii (BB)	29, 0, M	H29MB0	H3
	Biceps brachii (BB)	27, 0, F	H27MF0	H4

‡ Muscle Age, duration (Years), sex (M/F)

(i.e., $j = 1, 2, 3, 4$). Then, the correlation between the consecutive features (i.e., $X \in \mathbb{R}^{10}$) is evaluated. Estimated correlations among their pairs {1,2}, {3,4} and {5,6} are presented in γ_{IJ}^{j}, which is referred to as CCCM (i.e., Eq.(3.25)). The notations I and J represent group and subgroup respectively. Each row $\gamma_{IJ}^{j}(m.,)$ denotes the correlation of consecutive features. However, due to space constraint, in each case only upto five elements of correlations are shown in CCCMs in our case studies. More specifically, the variation in correlations are more pronounced in graphical analysis, where correlation is plotted against the projection dimension (i.e., d).

$$CCCM = \gamma_{IJ}^{j} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1d} & r_{1i} \\ r_{21} & r_{22} & \dots & r_{2d} & r_{1i} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{md} & r_{mn} \end{pmatrix} \tag{3.25}$$

ALS: Variations among the features of $A1$ in terms of correlation for four random observations is shown in Fig. 3.10(a). In addition to that, quantitative measures presented in CCCM also provide better understanding of these variations. Each plot corresponding to each row of the CCCM indicates the correlation of input features pairs extracted from specific signals. Consecutive rows of the CCCM denote the correlations among the consecutive feature pairs. Indeed, it is suitable and flexible way to represent underlying relationships. It is seen that $\gamma_{IJ}^{j}(1.,) \neq \gamma_{IJ}^{j}(2.,) \neq \gamma_{IJ}{}^{j}(3.,)$ which reveals the dissimilarities of features. It is obvious in many real-world phenomena, however, beyond three to four columns, it is minimum with few exceptions. In other words, transformed feature variation corresponding to lower order correlation are insignificant. Furthermore, dissimilarities among the features conclude that various feature sets, even extracted from same signal, share different nature and quality information. Therefore,

there is an obvious challenge to devise suitable strategy for more elaborate solutions.

As is evident that the variations of signals within same subgroup are also prominent (Fig. 3.10(a) and therein). Fig. 3.10(a)-(c) depicts similar statistical analysis in A2-A3. Accordingly, the nature of transformed FVs will be different. Nonetheless, the FVs corresponding to higher order correlation are more significant and overlapping while lower order FVs deviating in nature and carry more orthogonal behavior as shown in Fig. 3.10(d). For example, transformed feature-pairs $[U_1, V_1]$ and $[U_2, V_2]$ of pair {3,4}, $j = 2$ with $r = 0.99, 0.50$ show coherent in nature, while other pair $[U_3, V_3]$ with low correlation $r = 0.25$ shows orthogonality (non-overlapping) character. This evinces that the features with low correlation are less informative for correlation based analysis. Furthermore, the stability of input feature-pairs can be examined from the closeness of their correlation profiles. Low deviation between any two correlation profiles indicate high stability of input feature-pairs and stationary nature of signals e.g., the feature pairs-{3,4} and {5,6} w.r.t. {1,2} are more stable as indicated by the rows $\gamma_{11}^{1}(2.,)$ and $\gamma_{11}^{1}(3.,)$ in $\gamma_{11}^{j=1}$. (Eq.(3.26)). This analysis provides sharing information contents among the features and thus, it helps to select appropriate features dimensionality. However, it does not account the uncorrelated information between any two features. Experiment has been carried out on more number of signals (i.e., $j = 4$ to 10) under same age-group with common disease profile. However, no significant exception is seen. Besides aforesaid estimations, the OCs are also evaluated using Eq.(3.10), which are found as $[1.2, 1.3, 1.3]$, $[1.5, 1.5, 2.0]$, $[1.3, 1.3, 1.5]$ and $[1.5, 1.3, 2.6]$ for j=1, 2, 3, 4. Low values of OCs are mainly due to similarities of lower order transformed features. Even though, mean OC demonstrates significant extent of intra and inter signal variation.

$$\gamma_{11}^{j=1} = \begin{bmatrix} 0.98 & 0.28 & 0.20 & 0.18 & 0.14 \\ 0.99 & 0.50 & 0.25 & 0.17 & 0.15 \\ 0.98 & 0.41 & 0.30 & 0.18 & 0.15 \end{bmatrix}, \gamma_{11}^{j=2} = \begin{bmatrix} 0.97 & 0.46 & 0.44 & 0.40 & 0.30 \\ 0.98 & 0.69 & 0.56 & 0.54 & 0.42 \\ 0.98 & 0.61 & 0.59 & 0.48 & 0.44 \end{bmatrix} \tag{3.26}$$

Fig. 3.10(b) and Fig. 3.10(c) depict similar analysis in A2 and A3 respectively. In A2, minimum variations for $j = 2, 3, 4$ and prominent variation for $j = 3$ are observed. The OCs over all observations-$[0.98, 0.58, 0.53]$, $[0.97, 0.55, 0.46]$, $[0.98, 0.60, 0.46]$ and $[0.98, 0.60, 0.36]$ are approximately equal. Furthermore, the mean distributions remains approximately constant under same subgroup irrespective of j. Thus, it reveals that signals are highly stable and more or less carries equal energy content. In such case, inclusion multiple signals for feature extraction enhances learning parameters. Although, the statistics for A2 and A3 are similar, they differ by recording sites, i.e., A2 and A3 signals were collected from muscle *TA* and muscle *VL* respectively. It indicates that consistency might not depend on sites but on disease profile. The fluctuations in A4 are

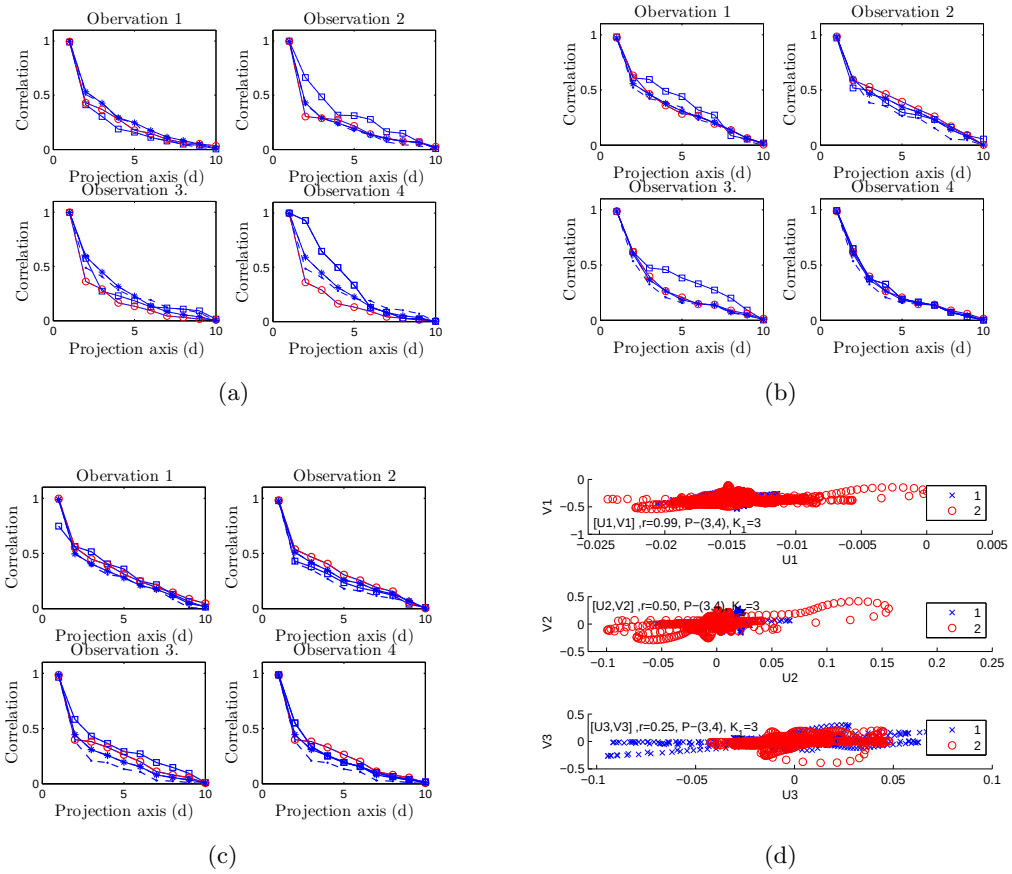

Fig. 3-10: Graphical representation of correlation profiles in CCCM for $j = 1, .., 4$ obtained correlation analysis in (a) A1, (b) A2 and (c) A3 of ALS. (d) indicates clusters formed by transformed features $(\{U_1, V_1\}, \ldots, \{U_3, V_3\})$ of input-feature pair$\{3,4\}$ with $j = 2$. (Legend for (a)-(c) 1st pair-□; 2nd pair-⋆; 3rd pair-· and mean-◯).

higher than that in A2 and A3 and less than that of A1. Further investigation on different age subgroup with common disease profile confirms the fact. Thus, in ALS, under mild or less severe cases, signal fluctuation from frame to frame (i.e., feature to feature) is more prominent than severe cases irrespective of gender levels. The OCs on A4, A2 and A3 are $[1.5, 1.9, 2.1]$, $[1.7, 1.9, 2.4]$, $[1.4, 1.8, 2.1]$, and $[1.5, 1.7, 1.9]$; $[1.7, 1.8, 2.3]$, $[1.8, 1.9, 1.8]$, $[1.7, 1.7, 2.1]$, $[1.5, 1.6, 1.7]$; and $[2.1, 1.9, 2.6]$, $[1.6, 1.8, 1.5]$, $[1.4, 1.8, 1.8]$, $[1.2, 1.5, 1.5]$ respectively, which also indication of extent of variations.

Myopathy: Fig. 3-11 shows variations in myopathy groups. It can be seen that the variation in r among the features in these subgroups are smaller than that of ALS. It is presumably due to significant dissimilarities among the features of two dissimilar groups. A close inspection reveals that variations in r is significant only in profile with higher disease duration. Mean r of M1 and M2 in Fig. 3.11(a) and Fig. 3.11(b) further corroborate the visual findings.

Normal: Fig. 3.12(a) and Fig. 3.12(d) evince that normal subgroup of aged 23 and 37 (i.e., H1 and H4) show significant variations, whereas it is similar in subgroups of aged 26 and 29 aged (i.e., H2 and H3). In addition to that, the estimated OCs also

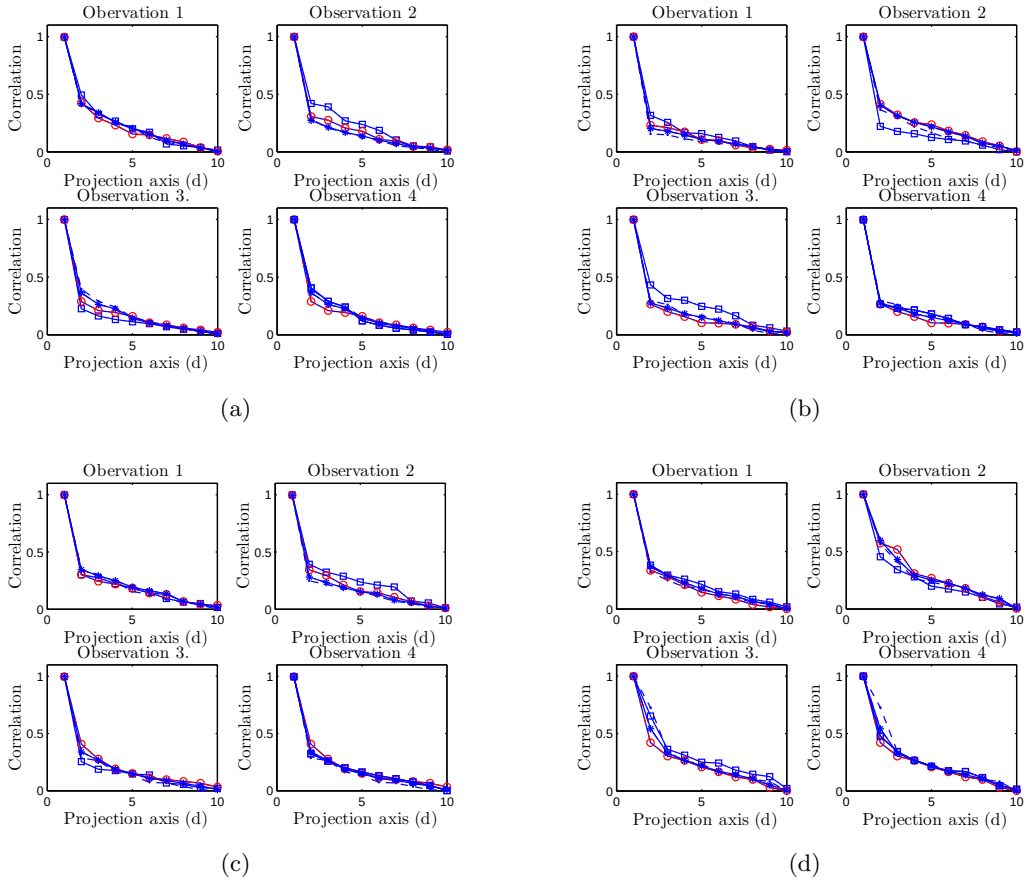

Fig. 3-11: Graphical representation of correlation profiles in CCCM for $j = 1, .., 4$ obtained correlation analysis in (a) M1, (b) M2, (c) M3 and (d) M4 of myopathy.

indicate the same. It is to be noted that the nature of variations in H4 are similar with close aged-H2. Thus, the variations within signals are significant toward two extreme age of subjects, but in case of the middle age, it is not so prominent. Thus, it infers that signal variations of mayopathy and normal subgroups are symmetrical while ALS subgroups are asymmetrical irrespective of disease duration. Further, it indicates the dependency on the type of disease.

3.5.1.2 Variability in intra-subjective signals (VIS)

VIS is a technique to investigate how signals of one subgroup varies with signals of other subgroups under the same group. It ensures whether it is necessary to consider multiple signals of various subgroups or randomly selected signals for feature extraction or not. Although such variations are obvious in many real-time of biomdical signals, VIS can be used to properly utilize the available resource for evaluation of usable feature.

To implement VIS, the feature variables S_1 (=4) of one age-matched are correlated with corresponding variables of other subcategories under same disease group (e.g., ALS). Therefore, VIS is evaluated between corresponding pairs, as indicated by

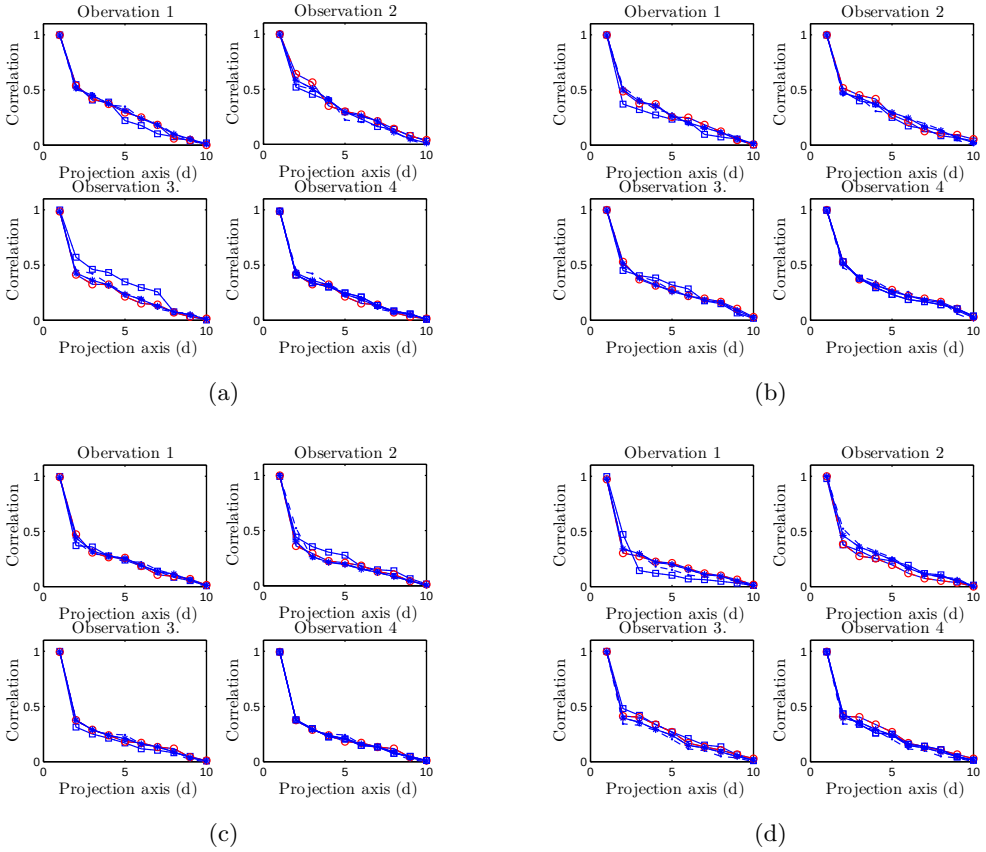

Fig. 3-12: Graphical representation of correlation profiles in CCCM for $j = 1, .., 4$ obtained correlation analysis in (a) H1, (b) H2, (c) H3, and (d) H4 of healthy controlled group or normal.

$r_{11}, \ldots, r_{44}$ in Table 3.3. The parameter μ_r indicates the mean correlation of four measurements (i.e., horizontal mean), which is assumed as ideal reference to establish the global variation. For example, between-subgroup feature variables are annotated with code such as A_{25} which indicates features of ALS subject 2 of age 35 and subject 5 of age 52. Table 3.3 show the variations among various possible combinations of features which demonstrate the inter-feature variations. However, a close inspection reveals that these variations are almost constant except in few cases. In ALS dissimilarities are seen over features (i.e., signal frames), specifically in first and second rows (i.e., A_{25} and A_{56}). In case of equivalent age range, it is insignificant. Furthermore, it is evident that the mean values are also approximately constant throughout the analysis. Similarly, small variations are also observed in myopathy and normal groups. It is prominent from r_{22} in both cases. Thus, inter-subjective signals of myopathy and normal groups are more stable than that of ALS groups.

3.5.1.3 Feature extraction, reduction and statistical validation

Selection of order of feature is one of the most important tasks in machine learning applications. In data fusion model, specifically in high dimensional analysis, data reduction

Table 3.3: VIS measurement in inter-subjective feature pairs in terms of pair correlations r_{ii}, where i features.

Subject	r_{11}	r_{22}	r_{33}	r_{44}	μ_r
A25	0.98,0.46,0.37,0.25	0.98,0.47,0.40,0.29	0.98,0.67,0.49,0.45	0.99,0.54,0.48,0.38	0.98,0.54,0.44,0.34
A56	0.99,0.45,0.27,0.26	0.99,0.53,0.48,0.36	0.99,0.72,0.54,0.44	0.99,0.72,0.49,0.45	0.99,0.60,0.44,0.38
A67	0.98,0.32,0.29,0.26	0.99,0.55,0.36,0.34	0.99,0.40,0.28,0.23	0.99,0.35,0.33,0.24	0.99,0.40,0.31,0.27
A37	0.98,0.35,0.27,0.22	0.98,0.33,0.25,0.17	0.99,0.31,0.30,0.29	0.99,0.33,0.29,0.28	0.98,0.33,0.28,0.24
A34	0.99,0.39,0.32,0.29	0.99,0.38,0.32,0.29	0.99,0.46,0.38,0.28	0.99,0.44,0.36,0.28	0.99,0.42,0.35,0.29
M14	0.99,0.23,0.23,0.17	0.99,0.26,0.22,0.18	0.99,0.27,0.22,0.21	0.99,0.32,0.25,0.21	0.99,0.27,0.23,0.19
M16	0.99,0.25,0.20,0.16	0.99,0.29,0.26,0.20	0.99,0.32,0.27,0.18	0.99,0.25,0.20,0.13	0.99,0.28,0.23,0.17
M23	0.99,0.32,0.25,0.23	0.99,0.32,0.31,0.24	0.99,0.33,0.23,0.20	0.99,0.29,0.25,0.21	0.99,0.32,0.26,0.22
M25	0.99,0.47,0.32,0.20	0.99,0.30,0.26,0.24	0.99,0.38,0.29,0.24	0.99,0.35,0.30,0.28	0.99,0.37,0.29,0.24
N56	0.99,0.48,0.39,0.25	0.99,0.39,0.27,0.22	0.99,0.47,0.37,0.28	0.99,0.47,0.35,0.27	0.99,0.45,0.35,0.26
N42	0.99,0.39,0.35,0.33	0.99,0.52,0.37,0.34	0.99,0.53,0.49,0.42	0.99,0.50,0.46,0.39	0.99,0.49,0.42,0.37
N41	0.99,0.37,0.33,0.30	0.99,0.48,0.32,0.30	0.99,0.37,0.33,0.26	0.99,0.36,0.34,0.27	0.99,0.40,0.33,0.28
N13	0.99,0.34,0.26,0.24	0.99,0.34,0.28,0.24	0.99,0.37,0.33,0.28	0.99,0.31,0.28,0.25	0.99,0.34,0.29,0.26
μ	0.99 0.37 0.30 0.24	0.99 0.40 0.32 0.26	0.99 0.43 0.35 0.29	0.99 0.40 0.34 0.28	0.99 0.40 0.32 0.27
σ^2	0.0 0.01 0.003 0.002	0.00 0.01 0.012 0.04	0.004 0.02 0.01 0.01	0.0 0.02 0.01 0.01	0.0 0.01 0.01 1 0.004

A25-ALS(35-52), A56-ALS(52-56), A67-ALS(56-60), A37-ALS(60-61), A34-ALS(61-67), M14-MYO(26-28), M16-MYO(28-33), M23-MYO(41-44), M25-MYO(44-63), N56-Nor(21-23), N42-Nor(26-27), N41-Nor(27-29), N13-Nor(29-37).

step is essential preprocessing step to avoid over-fitting of model as well as to form of compression [125]. Focusing on open issue of choosing an appropriate order of feature, VWS and VIS are addressed intending to extract latent information across the datasets. As is evident, intuitively first few sets of features have significant proximate behavior and can efficiently capture the underlying information from input variables. These low dimensional feature preserve most of the energy contents in the form of statistical measures which well-facilitate learning tasks.

This investigation provides an efficient solution that maximally reduces the dimensionality and also retains the joint information. Following the feature-pair correlation, the local feature dimensionality can be reduced based on correlation index and subsequently evaluated collective features of all possible feature pairs using central processing model would depict more meaningful measure. Based on aforementioned analysis, four different combination of features (i.e., Type of feature) are derived for subsequent investigations (See, Table 3.4). Type-I includes d-dimensional m pairs (i.e., m=k/2=3) features as a low dimensional view. For non-stationary bio-medical signals, this type of feature depicts proper measure, despite having symmetry among various time-frames or its feature representation upto a certain extent. Type-II includes R-domain (i.e., 2) d-dimensional m pairs feature. It represents multiple replica of type-I feature, i.e., combination of direct and DWT feature spaces, that improves quality of obtained feature space. In case of high variations among signals, type-III retains more information about the process, which includes S_1 (i.e., 4) signals. In other words, type-III is the combination of S_1 number of type-II feature. For multiple modality data analysis [13], type-IV would carry more physiological information. However, our study does not address such analysis. The aforementioned joint feature representations are intended to avoid the most common shearing information which unnecessarily increases learning parameters. Proper choice of feature representation is essential both at local and central fusions in order to avoid dimensionality issue.

Feature fusion enables providing a clear picture that capable of exploiting the underlying structure of heterogeneous features and allow better interpretability. Al-

Table 3.4: Proposed various combination of features along with the nature of signals.

Feature type	Pattern	Nature of profile [Symmetry/Asymmetry](Y/N)
Type-I	$d \times m$	Y (e.g.,Fig. 3.11(c), Fig. 3.12(c) for $j = 1, 3, 4$)
Type-II	$R\times\ d\times\ m$	N (e.g., Fig. 3.10(a), Fig. 3.10(c))
Type-III	$S_1\times\ R\times\ d\times\ m$	N (e.g., Fig. 3.10(a) for j=2, 4)
Type-IV	$N\times\ S_1\times\ R\times\ d\times\ m$	N (For multi-modalities data such as [13])

though all types of feature fusion eventually yield single feature frame, some relationships are more suitable in the context of dimensionality and also allow to cover wide range information. Therefore, this study focus on evaluating single form of discriminant features using first three feature representations stated in Table 3.4 with the help of model Eq.(3.12a)-Eq.(3.13b) and Eq.(3.14a)-Eq.(3.14d). It is admitted that different features, i.e., feature types are suitable for different input environments which have been investigated later.

Concatenation or summation are two beneficial feature fusion techniques [5]. However, summation fusion is more efficacious due to dimensionality concern in comparison to the concatenation. Thus, DMII with F=1,2 is more superior to DMI. The local CCDFs are estimated and then fused them through global fusion scheme. For high dimensional case, i.e., mean of all local features are summarized as global features (i.e., DMI and DMII) instead of juxtaposing them similar to Eq.(3.14a). Feature space is then subjected to one-way analysis of variance (ANOVA) test which is performed in MATLAB at 95% confidence level. It ensures whether the selected feature spaces are statistically significant or not [133]. Following the analysis, 10-dimensional feature matrices are chosen. In this analysis, any feature having $p > 0.05$ is considered insignificant and removed from the feature space. However, it is worth noting that feature vectors are standardized before fusion to avoid the diverge nature even extracted from the same pattern [5, 32]. If $A^1,\ A^2, .., A^d$ are feature vectors, then features are normalized by $\hat{A^i_j}$=$A^i_j - E[A^i_j]/\sigma$,$j = 1, .., d$, A^i_j is the j^{th} component of i^{th} feature so that the normalized feature has $E[\hat{A^i_j}]$=0 and variance i.e σ^2=1 [134]. The effectiveness of the proposed feature generation scheme is further corroborated by feature distribution as shown in Fig. 3-13 and separability analysis as outlined in Table 3.5. Separability measures (SM) among feature groups are estimated using two Fisher criteria that do not require the Gaussian [135].

$$J_1 = \frac{trac(S_B)}{trac(S_W)}, \tag{3.27a}$$

$$J_2 = trac(S_B^{-1} S_W). \tag{3.27b}$$

where S_W and $trac = \sum a_{ij}$, $i = j$ (i.e., sum of diagonal elements), a_{ij} matrix elements, are represent within-scatter matrix and trace of square matrix respectively. Larger values of J_1 and J_2 indicate well-separation of features in decision surface. Thus, it ensures minimum possibility of getting poor performance while applying the models for

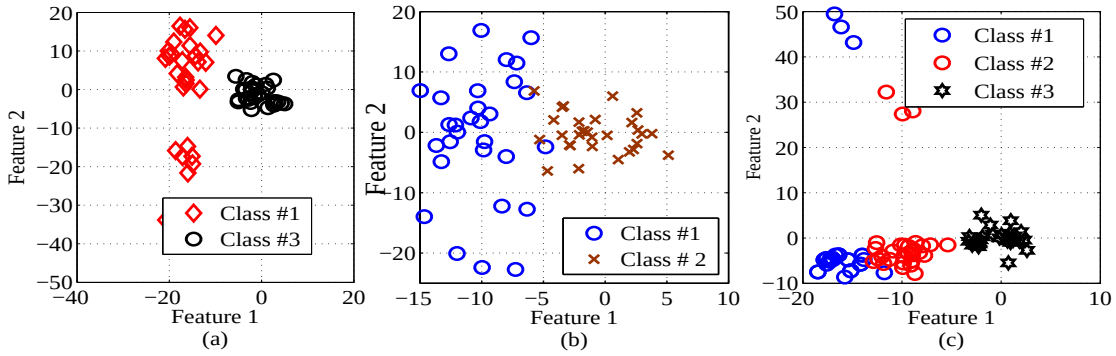

Fig. 3-13: Scatter feature distribution using LDA, a) two-class and b) three-class under scheme Type-II (Here #1-ALS, #2-Myopathy and #3-Normal).

Table 3.5: Comparison of estimated J_1 and J_2 values with few reported methods. Highest values are indicated by boldface.

Methods	J_1	J_2
Conventional PSD based methods [136]	0.69	1.82
STFT and RVM [136]	0.74	4.57
Q-factor wavelet and spectral features [137]	2.0739	148.4636
Proposed feature model	**3.987**	**149.0125**

subsequent task. The SM in Table 3.5 shows that the features generated via fusion strategy are well-separated and it passes the statistical null-hypothesis that have low-dispersed scatter distribution.

3.5.1.4 Performance analysis of feature fusion based learning

Features Type-I, Type-II and Type-III are fused using DMI and DMII and obtained global feature discriminant descriptors are further subjected to statistical models. The significant features are embedded to learning model. The simplest k-nearest neighbors (k-nn) which requires only tuning parameter k, is taken for inferences. Choice of k-nn is due to wide popularity, ease of use and learning [58]. Optimum value of k is obtained using inter-cross validation technique.

To evaluate the performance of fusion based data-driven models using cross-validation technique, the dataset EMG_{N2001} is divided into three subsets as follows:

i). 2-class normal-ALS and normal-myopathy small dataset with 15 ALS, 15 normal and 15 myopathy recordings.

ii). 2-class normal-ALS and normal-myopathy larger dataset with 50 ALS, 50 normal and 50 myopathy recordings

iii). 3-class normal-ALS-myopathy larger database with 50 ALS, 150 normal and 50 myopathy recordings.

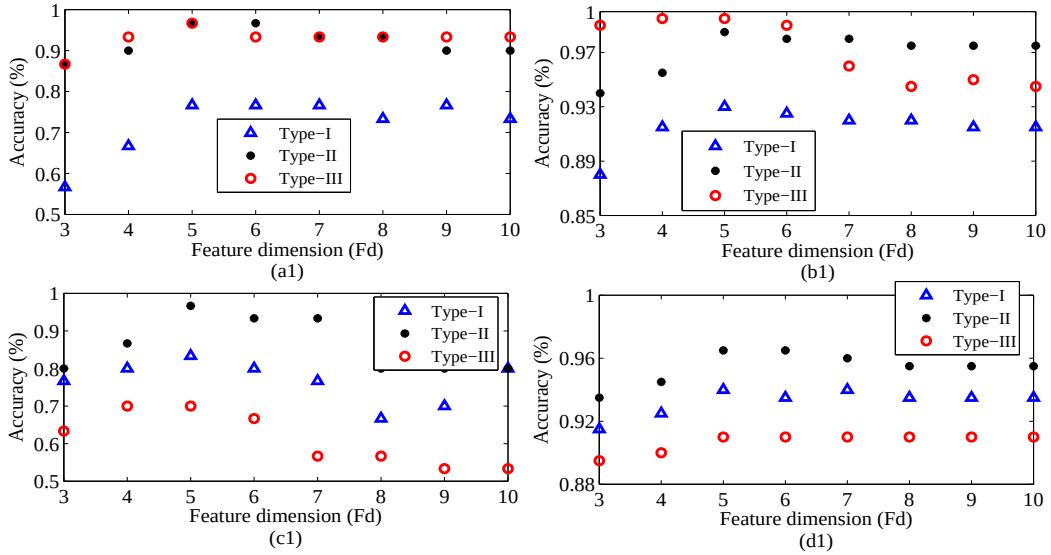

Fig. 3-14: Algorithm performance in terms of accuracy with different choices of features in binary two normal-ALS (a1-b1) and two normal-myopathy subject groups (c1-d1).

This partition is made similar to [58] for better comparison with this method and methods therein, and reasonable justifications. Additionally, the proposed method is also validated using another dataset EMG_{GNRC} collected from GNRC hospital, Assam, India during the period of 2014-2015.

Full scale model performance is first presented so as to give overview of changes in performance with feature dimension. However, Table 3.6 shows results at optimum feature dimension. Fig. 3-14 depicts the classification accuracy Ac of our algorithm. It is seen that accuracies attain optimal level for a particular feature dimension (Fd) and then gradually fall with Fd. In small dataset as in Fig. 3-14(a1), Type-II and Type-III show similar recognition rate whereas Type-III provides an optimal recognition rate as shown in Fig. 3-14(b1). It indicates that Type-III is more suitable in capturing information from highly variational signal profiles such as in ALS. The algorithm with Type-II yields good recognition rate in two binary datasets (i.e., normal and myopathy) as shown in Fig. 3-14(c1) and 3-14(d1). Thus, it ensures the robustness of the proposed scheme in small variation of profile. Nonetheless, in both cases, the algorithm with Type-I yields low Ac which is mainly due to elimination of orthogonal features which is affirmed from higher performances of Type-II and III features.

Fig. 3-15(a2)-(b2) show mean Ac in normal-ALS and normal-myopathy datasets, and Fig. 3-15(c2) depicts the Ac in three-class dataset. It can be seen that the method achieves optimal accuracy with type-II feature in three-class large dataset, which is due to low order multi-domain discriminant features. Statistical paired t-test that indicates significant group difference between ALS-myopathy and healthy control groups, also supports the proposed model. In addition, the algorithm maintains the consistency in performance with low deviation. It, thus, ensures the robustness of proposed technique and it also suggests to avoid unnecessary features that do not contribute in

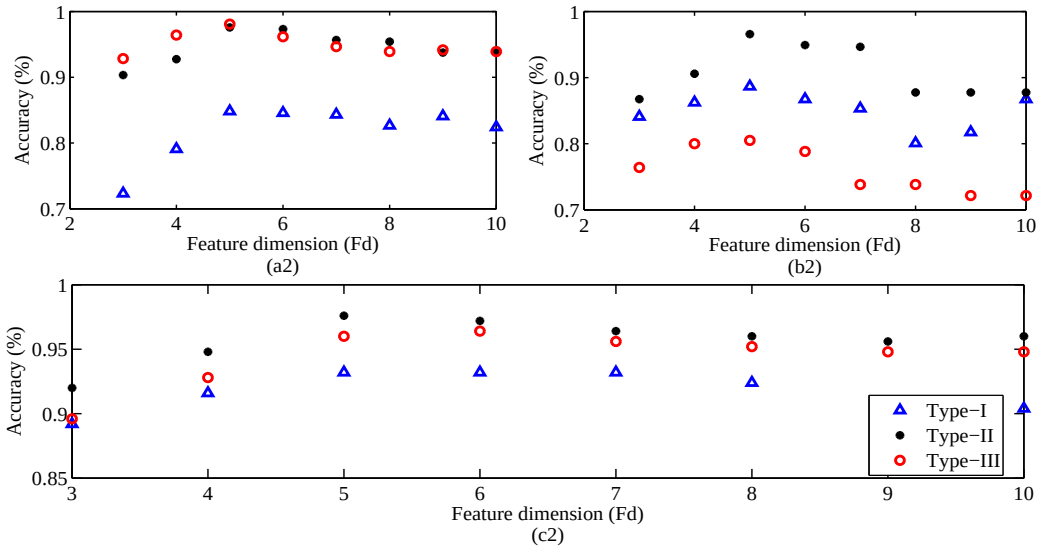

Fig. 3-15: Algorithm performance in terms of mean accuracy (two normal-ALS group (a2); two normal-myopathy subject group (b2)) and accuracy in three-class subject group.

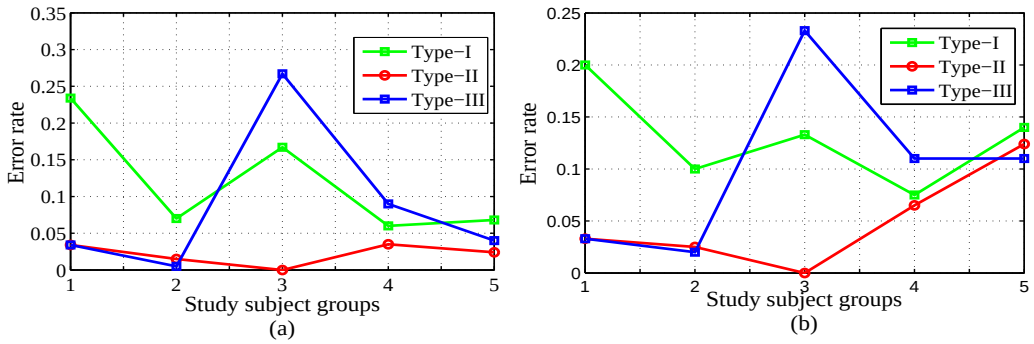

Fig. 3-16: Error rate of (a) DMI and (b) DMII in five groups of subject with three different features.

enhancing the performance.

Apart from quantitative markers as outlined in Table 3.6, PPVs on two smaller groups are $[0.87, 0.97, 0.99]$ and $[0.86, 0.92, 0.81]$ respectively and it is $[0.96, 0.97, 0.98]$ on three-class group. These parameter increase up to 2.6% (Type-II) and 8.8% (Type-III) in combined analysis. Similar inferences can be made from estimating NPVs as well. It also indicates the superiority of Type-III and Type-II in small and large-class datasets. It is to be noted that DMI with regularization and non-regularization provides quite similar outputs in small dataset. However, in three-class large datset, it shows degraded performance (i.e., 66.6% in Type-I, 83.3% in Type-II and 86.6% in Type-III). Fig. 3-16(a) and Fig. 3-16(b) depict the recognition errors in DMI and DMII. Higher level of error in DMII is presumably due to high feature dimensionality of DMII. These are 7.6% (DMI) and 9.3% (DMII) in small and large datasets respectively.

Table 3.6 shows the performance of our method as well as comparison with relevant reported methods. It is to be mentioned that the results of our methods with Type-II and III features are mainly compared with others since Type-I feature does not

Table 3.6: Performence measurements in terms of percentage (%) and comparison with state-of-art EMG methods.

Method	Type of feature	Study Group	Sn	Sp	Ac
Direct	DWT Max [138]	[A,N]	87.00	98.30	95.50
Direct	DWT ten Max [138]	[A,N]	87.00	98.60	95.80
Direct	AR+DWT [67]	[A,N]	78.90	80.20	79.40
MUAP	Morphological [64]	[A,N]	48.65	97.00	79.10
dMUAP	DWT [58]	[A,N]	99.00	100	99.70
Our method^{A}	Type-I	[A,N]	88.30	77.50	84.80
	Type – II	[A,N]	**96.50**	**99.00**	**97.50**
	Type – III	[A,N]	**99.65**	**96.65**	**98.10**
	Type-I	[M,N]	91.30	80.00	88.65
	Type – II	[M,N]	94.00	94.95	**96.50**
	Type-III	[M,N]	83.30	74.30	80.50
Direct	DWT Max [138]	[A,M,N]	86.00^{a}, 87.70^{b}	82.70	87.60
Direct	DWT ten Max [138]	[A,M,N]	87.00^{a}, 88.30^{b}	89.00	89.40
Direct	AR+DWT [67]	[A,M,N]	60.90^{a}, 70.00^{b}	74.00	72.40
MUAP	Morphological [64]	[A,M,N]	53.50^{a},63.00^{b}	83.00	71.40
dMUAP	DWT [58]	[A,M,N]	99.00^{a}, 98.00^{b}	100	99.50
Our method$^{\mathbf{A}}$	Type-I	[A,M,N]	96.6^{a},90.0^{b}	84.0	93.2
	Type – II	[A,M,N]	92.00^{a},**96.00$^{\mathbf{b}}$**	**98.60**	**97.60**
	Type – III	[A,M,N]	**100$^{\mathbf{a}}$**,**82.00$^{\mathbf{b}}$**	**99.30**	**96.00**
Our method$^{\mathbf{B}}$	Type-II	[A,M,N]	100^{a}, 100^{b}	100	100

Note:-^{a}Sensitivity of ALS (SnA), ^{b}Sensitivity of Myopathy (SnM), $^{\mathbf{A}}$-EMG$_{N2001}$ and $^{\mathbf{B}}$-EMG$_{GNRC}$.

include statistically orthogonal features. For better understanding of the efficacy and superiority, the results for three features are outlined herein.

A close inspection of Table 3.6 manifests that proposed learning provides promising performances which is also closer to [58]. Most of the cases, the proposed method shows superior performances. Integrity of advocated method is due to well-defined feature extraction and fusion strategy. It captured unique features that decreases the learning complexity which in turn improves the recognition rates maintaining the consistency over various datasets. High performance is presumably due to incorporation of orthogonal features into global feature descriptors. In obtaining high performance, signal decomposition strategy for formulating features plays a vital roles. Moderate level of signal decomposition is more significant. Too small or too large signal decomposition fail in maintaining symmetry between consecutive pair of feature variables. In that case, CCA fails in extracting valuable information. It is ensured by using EMGLAB and MATLAB. Or method requires 38.485 s, 42.346 s and 71.814 s (mean over three runs) for classification of two-class (small and large) and three-class problems, including feature extraction. However, these are machine specific and they may vary with classifier and machines. The performances over EMG$_{GNRC}$ that includes 60 recordings (20 recording/subject group), are 100% (Ac) and 100% (Sn and Sp). This ensures the consistency of algorithm over wide variety of real-time data.

3.5.1.5 Comparison analysis: MVF-Case-I

This section briefly describes the performance of proposed method and compares with various state-of-art EMG methods outlined by Doulah *et al.* [58]. Table 3.6 shows results over two binary groups [ALS, myopathy], [myopathy, normal] and three-class subject group [ALS, myopathy, normal] which are represented by [A, M], [M, N] and [A, M, N]. It further includes type of feature and classifiers used by various methods.

The proposed method surpasses most of the previously reported methods except dominated MUAP (dMUAP). However, dMUAP method is often tricky and requires the guidance of domain experts. The adopted method achieves much higher accuracy (i.e., 97.50%, 98.10%, 97.60%), sensitivity (i.e., 99.65%, 100^{a}%) and specificity (i.e., 99.00%, 98.60%) than the methods in [64, 67, 138]. Two sensitivities-SnA and SnM for two positive cases-ALS and myopathy are evaluated separately, which show higher values than the method in [58]. It is observed that our approach achieves optimal accuracy under Type-III feature pattern in binary class and under type-II in three level class. However, Type-III involves number of observations and hence, it requires comparatively more processing time, therefore Type-II is more suitable and promising. In achieving usable information from large data, the advocated model provides the well-defined framework that leads to optimal performance with minimum variances. It, thus, evinces the integrity of this learning framework.

It is worth mentioning that direct methods-DWT max, DWT ten max [138] and MUAP methods-morphology, dMUAP [58, 64] utilize discriminant feature vectors from each pre-defined signals. MUAP-based method [58] finds dMUAP which requires guidance knowledge of morphological patterns associated with diseases. However, MUAPs in ALS [54] are unstable and it is difficult to find the dMUAP. Futhermore, the nature of MUAP may vary due to a) sub-class of disease; b) nature of force to recruit the muscle; and c) clinical set-up [55, 83]. In such case, MUAP-based methods may not feasible for analysis. Despite good results over specific dataset, some methods are lack in providing consistent results over wide varieties of datasets. For example, DWT+AR based method achieves an accuracy of 95.0% in neuopathy, myopathy and normal dataset, while it is 71.4% in ALS, myopathy, normal which could be due to bias in feature registration or classifier. In contrast, our method is effective in utilizing multi-domain features using feature fusion techniques for promising and consistent results.

3.5.2 MVF-Case-II

This section first demonstrates the variations of correlation between intra-subgroup MVs outlined in Table 3.1 in Fig. 3-17. As is evident, low correlations indicate the significant dissimilarities of signals used to formulate the MVs. It is due to the fact that MUAP of EMG signal does not occur at same instances in different time-frames of signals.

Furthermore, subgroups MVs are differed by age. This measurement makes clear that signals in different subgroup of a given subject group are significantly different. Thus, it is reasonable to make use of best possible information of multiple MVs for the complete understanding of phenomena associated with study groups. Thus, it remains in favor of the MV-based scheme advocated in the section 3.4.

Unlike MVF-Case-I, MVF-Case-II formulates two strategies S-I and S-II using selection tree as in Fig. 3-7 in order to enhance more reliability and robustness of the proposed feature fusion based model that can easily overcome the theoretical bottlenecks of previous approaches.

3.5.2.1 Performance of the mCCA on real-time datasets

Fig. 3-18 shows wavelet components for optimal $q = 8$ which is obtained from performance and statistical test. This optimum number is taken from each subgroup-two from each subject ($= 2 \times 4$) and discriminant features are evaluated (See, section 3.4.4). One-way analysis of variance (ANOVA) is carried out at 95% confidence level and p=0.05 to assess the quality of features. Feature with $p > 0.05$ is discarded from feature matrices. Moreover, two-way ANOVA indicates that there is no benefit in adding more transformed feature while deriving discriminant vectors. A minimal drop of accuracy, at p=0.05 is observed while increasing feature dimension from 7 to 8. Features are subjected to LDA that finds optimum decision surface among the features. Table 3.7 and Fig. 3-19(b) show p-values of features and scatter distribution for S-II which show higher discrimination ability of features in comparison to S-I features.

In order to obtain the best feature combinations for S-I and S-II, performances are investigated on the training dataset of EMG_{N2001} in Table 3.8 over three repeated measurements. As is evident, the D, E (S-II) and E (S-I) (not shown) show optimal performance, which are (i.e., D (S-II) and E (S-I)) used for full-scale performance evaluation over three datasets independently using 3-fold cross-validations. It is seen that method achieves promising results specifically with S-II over two and combined datasets as shown in Fig. 3-20(a)-(c), whereas Fig. 3-20(d) depicts overall mean. However,

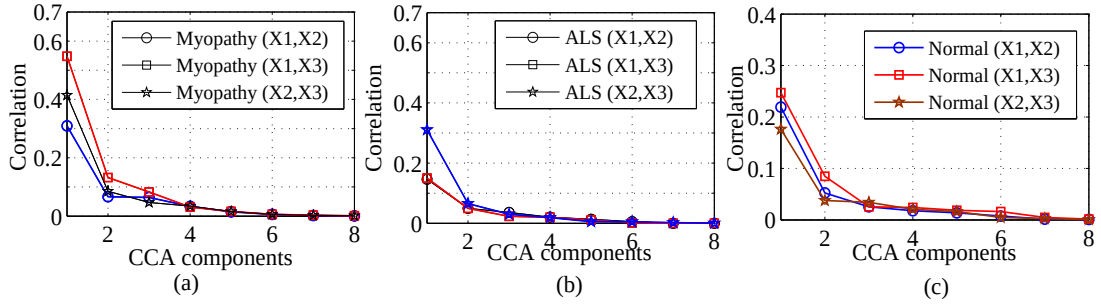

Fig. 3-17: Mean correlation among intra-subgroup MVs for three subject groups-a) myopathy, b) ALS and c) normal. Here X_1, X_2 and X_3 indicate within-subject group MVs as outlined in Table 3.1.

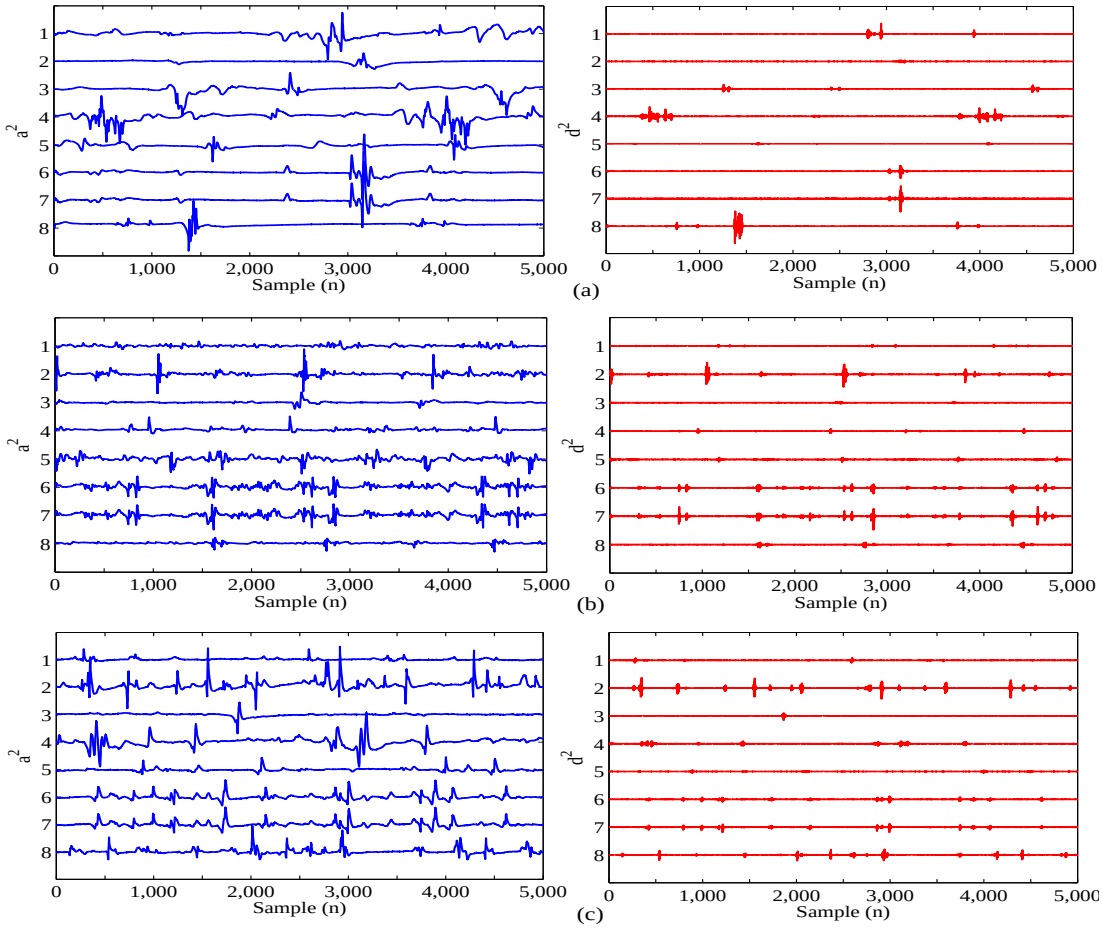

Fig. 3-18: Formulation of MV using DWT-coefficients approximate coefficients (a^2) of (a) ALS, (b) myopathy and (c) normal. Here d^2 are the second level detailed coefficients obtained during wavelet transformation. Unlike case I, this study employ multiple signals (i.e., q) in formulating the DWT-MVs.

results with S-I are also remarkable and comparatively better, especially, in EMG_{N2001}. SnA and SnM indicate the sensitivities for ALS and myopathy. With S-II, the algorithm shows the highest performance over EMG_{GNRC}. However, mean OA is 99.4%, and SnA, SnM and Sp are 99.5%, 97.7% and 100% respectively (See, Fig. 3-20(d) S-II). In all the cases, the classifier with S-II achieves promising results than that of S-I. Method provides good individual accuracy in categorizing normal subjects.

The confusion matrix over EMG_{N2001}+EMG_{GNRC} in Table 3.9 and Fig. 3-21 shows individual class predicting ability of model with S-II. Our model fails in categorizing only three cases (one ALS, two myopathies) while none of the control subjects is misclassified. However, in EMG_{GNRC}, the model accurately predicts the subjects (See, Fig. 3-20(c)[S-II]). Thus, the promising results and low variance in outcomes apparently indicate the reliability of the mCCA.

Results for $q = 8$ are ascertained from the performance assessment as shown in Fig. 3-22 and statistical test. ANOVA test at $p<0.001$ suggests not to add more than eight signals to avoid the computational burden. Ease of implementation and feature selection which enable adopting various classification models, are another vital

Table 3.7: p-values of selected features

Feature	p-value (S-I)	p-value (S-II)
f_1	$< 3 \times 10^{-3}$	$< 1 \times 10^{-3}$
f_2	$< 2 \times 10^{-3}$	$< 1 \times 10^{-3}$
f_3	1×10^{-3}	$< 2 \times 10^{-3}$
f_4	0	$< 2 \times 10^{-3}$
f_5	$< 3 \times 10^{-4}$	$< 2 \times 10^{-3}$
f_6	$< 4 \times 10^{-5}$	8×10^{-3}

Table 3.8: Classification performance of feature combinations of S-II over training dataset of the EMG_{N2001}

Combination	Feature	SnA (SnM)(%)	Sp (%)	OA (%)
A	f_1	88.0 (78.6)	94.8	87.9
B	f_1, f_2	88.0 (86.6)	96.9	93.1
C	f_1, f_2, f_3	93.3 (96.6)	96.6	95.5
D	**f_1, f_2, f_3, f_4**	**97.3 (97.3)**	**98.6**	**98.6**
E	f_1, f_2, f_3, f_4, f_5	**97.3 (97.3)**	**98.6**	**98.6**
F	f_1, f_2, f_3, f_4, f_5, f_6	93.3 (92.0)	98.2	95.5

consequences. It ascertains the benefits of MVs to extract relevant information from intramuscular EMG. However, upon increasing the signals beyond eight there is a drop of accuracy due to loss of useful data during the PCA. Use of more number of principal components might provide high recognition rate. However, it will increase the dimensionality of input spaces which in turn increases the complexity of feature. PCA provides good feature clustering assuming variance in data [20]. However, large input alters the feature clustering to a complex pattern which will degrade the performance of the PCA. In this case, training the system with an alternative technique could provide an efficient solution.

3.5.2.2 Reliability and scalability of the mCCA

The reliability of the mCCA is further explored in the context of the results obtained using discriminant analysis (DA) model with linear and quadratic discriminant functions [133]. Such extensive analysis affirms the reliability of the adopted feature extraction and fusion scheme since the specific model performance may not ensure the quality of the feature. As is evident, the feature statistic followed by the promising results in Fig. 3-20 ascertain the reliability of the scheme since the feature biasing is greatly reduced.

For scalibility measure, the features extracted using the datasets of EMG_{N2001} is used to estimate the performance over the EMG_{GNRC} as shown in Fig. 3-20(c)[S-II], wherein OA, SnA, SnM and Sp are 100%. Also, the performance is assessed over the EMG_{N2001}+EMG_{GNRC}. Fig. 3-23 shows the mean results over the datasets. It is

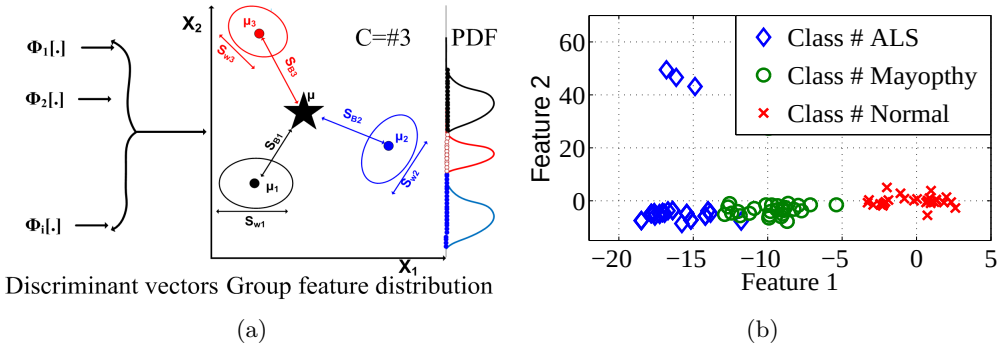

Fig. 3-19: a) Figure showing working principle of LDA for C-class problem (=3) with $C-1$ dimension, and b) feature distribution for S-II.

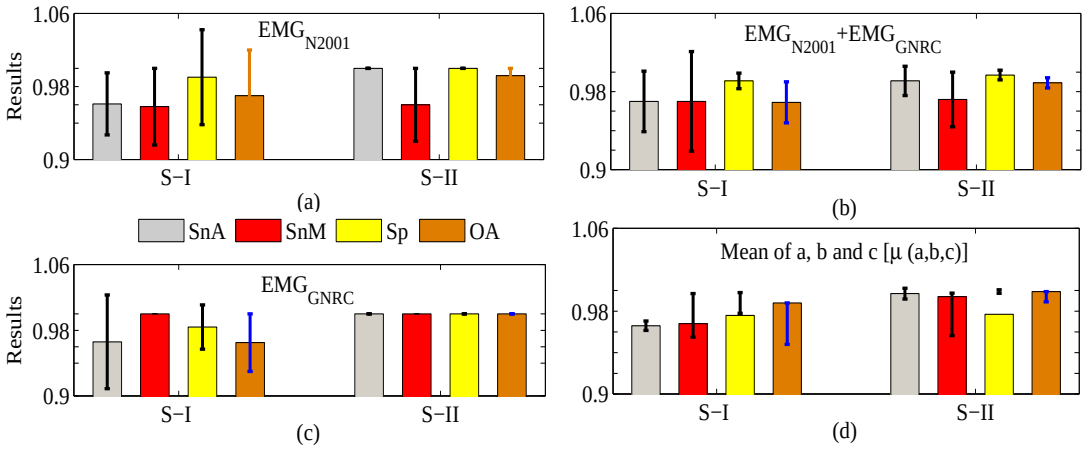

Fig. 3-20: (a)-(c): classification results of our method with D (S-II) and E (S-I) combinations over three datasets EMG_{N2001}, EMG_{N2001}+EMG_{GNRC} and EMG_{GNRC}, and (d): overall mean (i.e., of (a), (b), (c)).

further ascertained in the context of two data sets obtained from the EMG_{GNRC} itself. Although there are no common samples in the datasets, no changes in the results are observed ensuring the robustness and reliability of our proposed method.

3.5.2.3 Comparison Analysis: MVF-Case-II

This section also outlines the performances of various reported EMG methods and also focuses on the limitations, inherent feature structures in the context of our approach. In addition to that, the results of previously described strategy (i.e., MVF-Case-I) as shown in Table 3.6 are also highlighted to explore the improvement of the MVF-Case-II. This study focuses on direct and indirect comparisons. In direct-comparison, the results of our algorithm are directly compared with the results of previous methods, whereas in indirect-comparison relevant methods are implemented over our study dataset using common classifier (i.e., k-nn).

Table 3.6 depicts direct comparison analysis. Methods [70, 78] employed DWT

Table 3.9: Mean confusion matrix in terms of class-mean $\mu_{\mathbf{C}}$ (output classsification mean) and σ formulated from the results of the classifier with S-II. The diagnoal elements (boldface) indicate the correct classification

Actual	Predicted class-S-II					
	ALS		Myopathy		Normal	
	$\mu_{\mathbf{C}}$	σ	$\mu_{\mathbf{C}}$	σ	$\mu_{\mathbf{C}}$	σ
ALS	**69**	1	0	0	1	0
Myopathy	1	1	**68**	1	1	1
Normal	0	0	0	0	**170**	0

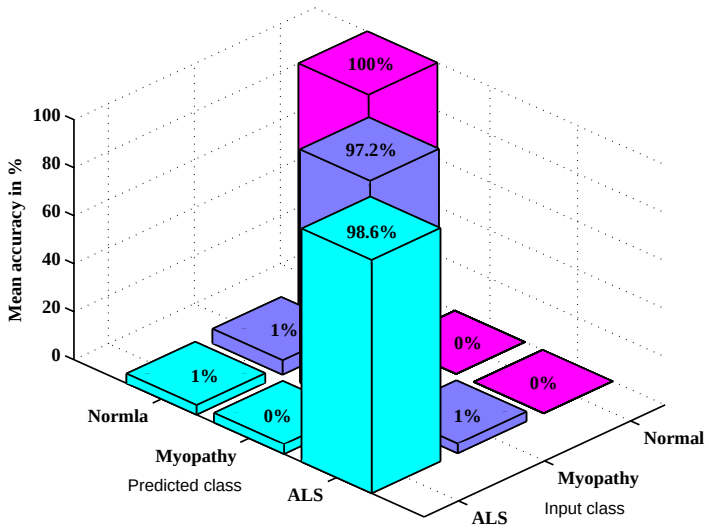

Fig. 3-21: Performance in terms of 3D confusion bars, wherein diagonal/off-diagonal elements indicate the correct/false prediction rate. Vertical axis indicates accuracy and horizontal axes represent actual and predicted outputs.

statistics for fuzzy-support vector machine (SVM) and particle swarm optimization (PSO)-SVM and investigated results over dataset of 27 subjects reported in [79]. However, low order statistics may not be feasible for highly nonlinear data. SVM requires optimal kernel parameters-C and γ for correct results and to ensure whether training samples of each class are uniform or not. However, in real-world, the effect of training samples are different. Methods [70,78] were derived from evolutionary algorithms which deal with binary problems. Computational burden and lack of theoretical guarantees are their inherent pitfalls. ESVM [79] selects kernel parameters using GA which involves a number of steps and have high computational time for larger datasets. RF [80] takes large memory and slow down the process in case of a large number of trees, lack of explanation and tends to overfit for noisy data.

Features AR+RMS+TD [74] often introduce high dimension and also take additional processing steps. AR+RMS and AR+RMS+TD give 28 and 44-dimensional feature respectively. Large mixture number incur very high computational expense but do not yield to good result for small data. Our study advocated the benefit of dis-

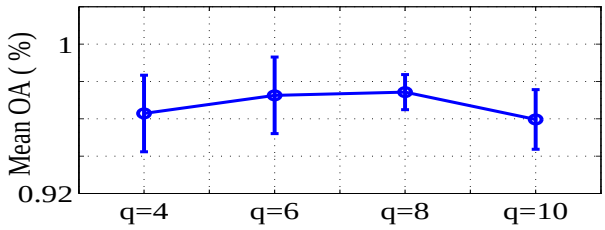

Fig. 3-22: Mean OA of S-I and S-II indicates the effectiveness of classification model for various choice of signals ($p = 0.0025, 0.074, 0.04$ and 0.065)

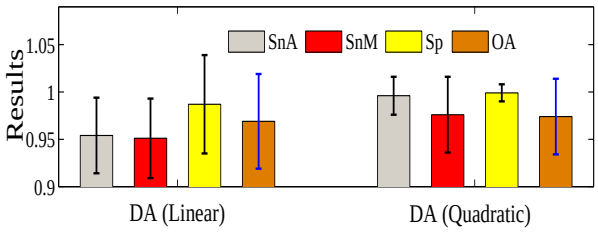

Fig. 3-23: Mean results with DA (linear) and DA (Quadratic) model. Performances are measured similar to the Fig. 3-20 using two different classifiers instead of k-nn.

criminant features for more reliable inferences. As is evident, the proposed method specifically S-II choice, outperforms many state-of-art methods and it also achieves an optimum accuracy of 100% in EMG_{GNRC}. Quantitatively, it improves recognition rate up to $2 - 2.6\%$ over the best value of reported method and it also maintains the consistency in performance.

Fig. 3-24 shows the results of direct comparison. Our results (i.e., P1-P3) are compared with implemented methods. Results of our preliminary small scale-MV study is shown in R5. Methods [70, 78, 79] and [80] employed 15, 23, 23 and 27 features (i.e., mean, average power, SD and ratio) evaluated from study signals. As is evident, our method is far superior to these methods. Reported methods show promising results over

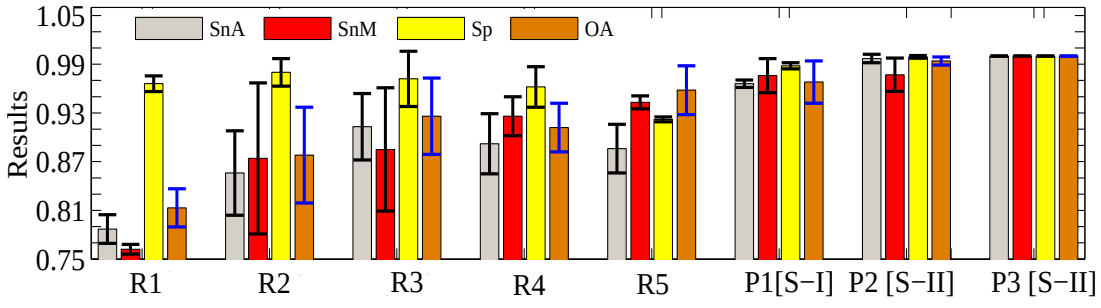

Fig. 3-24: Comparison of results in terms of SnA, SnM, Sp and OA (with $\pm\sigma$ as error) of various state-of-the-art methods, R1: Fuzzy-SVM [70], R2: PSO-SVM [78], R3: E-SVM, [79], R4: RF [80], R5: our preliminary small scale-MV with the proposed model as indicated by P1 [S-I] and P2 [S-II] for S-I and S-II respectively. Here P3 [S-II] indicates the performance of the proposed method with S-II over dataset EMG_{GNRC}. Although reported methods are recalled by their original nomenclature, but, all these methods employ k-nn as classification model instead of SVM, RF etc.

Table 3.10: Performance comparison with the state-of-the-art-methods. Boladface indicates the highest recogniation rate achieved by the proposed method.

Classification Model	Feature used	Study groups/subject	OA
SVM [76]	Frequency-feature	3/27	92.55
F-SVM [70]	DWT-statistical feature	3/27	93.50
ESVM [79]	DWT-statistical feature	2/27	97.00
PSO-SVM [78]	DWT-statistical feature	2/27	97.40
RF [80]	DWT-statistical feature	3/25	96.67
Proposed (S-I)	mCCA-statistical feature	3/37	96.8
Proposed (S-II)	**mCCA-statistical feature**	3/37	**99.4**
S-II [EMG$_{GNRC}$]	**mCCA-statistical feature**	3/12	**100**

small dataset. However, in EMG$_{N2001}$+EMG$_{GNRC}$, they show degraded performance. Our method improves recognition rate up to 3.6-4.2% over the results of reported methods.

Our method with S-II misclassified three cases which is presumably due to the inherent similarity of signals, order selection or dissimilarity of extracted features from the trained patterns. Integrity of advocated method is due to well-defined feature extraction and fusion strategy. It captured unique features that decreases the learning complexity which in turn improves the recognition rate. It requires average of 23.78 s and 36.46 s for feature fusion of includes S-I and S-II respectively. For classification task, it takes 2.1 s and 3.4 s. It is reasonable due to the incorporation of large information to have better diagnosis value.

3.6 Conclusion

This chapter addressed data-driven approaches based on feature fusion model using two independent multi-view feature generation schemes. In deriving decision making model for classification, we argued an efficient noise-feature dimensionality reduction technique based on variability and stability analysis of non-stationary EMG signals with an aid of multivariate statistical model. Then, a new feature extraction module was developed and subsequently a set of feature distributions were evaluated. The extracted feature patterns were fused using two proposed two global fusion techniques Thus, it was proved to be a useful benchmark for FES as a promising signal processing paradigm. Besides, statistical tests one-way and two-way ANOVA were performed to validate the feature space statistically. Finally, the proposed algorithm was investigated with two sets of data-publicly available and GNRC database, namely EMG$_{N2001}$ and EMG$_{GNRC}$. The outcomes of these investigations best fit with expert neurologists opinions. Finally, we ended this section with some consequences or advantages of the proposed method over previously reported methods, which are as follows:

i). It is a reversible process which makes the proposed scheme as a useful benchmark in reconstructing the signal after reduction of noise. Moreover, the scalability of the approach is quite high.

ii). Feature synchronization using two independent methods avoid loosing important information which are uncorelated between two set of feature. Hence, the possibility of missclassification rate is significantly reduced.

iii). The method provides quantitative evaluation of correlation between two sets of features which can be further extended to multiple sets of features. Thus, combined quantitative analysis and visual description of correlations among features help setting the threshold for reduction of feature dimensionality in order to avoid computational bottleneck of model. This strategy further reduces the complexity of learning process.

iv). Although its execution is time high for specific feature combination, but performance is promising which is the vital requirement for viable algorithm.

In the second approach (i.e., MVF-Case-II), we adopted a another multi-view feature generation scheme from EMG signal considering subjective age as one of important feature biasing factor. Accordingly, a subspace learning based feature fusion model which uses multi-view information extracted from EMG signal was developed. In this scheme, two multi-view feature formulation strategies, namely, S-I and S-II, were proposed and investigated the performance of the proposed model over various combinations of datasets EMG_{N2001}, EMG_{GNRC} and EMG_{N2001}+EMG_{GNRC}.

The MVF-Case-II technique employs multiple signals through MV feature generation tree using strategies S-I and S-II. It significantly reduces the feature dimensionality with an aid of optimization technique and the results suggest that information extracted from multi-view inputs yield an excellent mean recognition accuracy up to 99.4% with a specificity of 100% and sensitivities of 99.5% and 97.7%. Furthermore, the model achieved 100% performance over EMG_{GNRC} dataset. Admittedly, the classification efficiency relies on multi-view extracted from multiple signals belonging to various age-based subgroups of study subject groups. Furthermore, optimal signal selection for multiple views creation was investigated and validated using statistical analysis which revealed that no significant improvement in classification resulted from adding more than eight signals. The investigation shows the signals belonging to various subgroup according to age are significantly different. In view of such dissimilarities among the signals or extracted features the second approach is more suitable. Thus, the overall outcomes are encouraging and it envisages developing a graphical user interface which would be simple, accurate and reliable enough for clinical usage. It further promotes investigation with signals and development of reliable tool for quantitative decision support system. It is interesting to investigate how robust edition of CCA, namely, discriminant correlation analysis (DCA) play the role in enhancing the model performance and complexity by

deploying feature discrimination function. The proposed operational framework is not an alternative approach to replace the doctor in any way but can support and enhance the diagnosis process so as to take early alarm of patient health status for quick diagnosis. Thus, it promotes the implementation of such advance signal processing method in clinical and home care environments (i.e., portable devices based health monitoring) to improve the quality of life. The proceeding chapter will explore the extension of CCA (i.e., DCA) to further meliorate the algorithm performance and focus on reduction of theoretical bottlenecks for ease of implementation in various application domains.